

1-1-2011

Prestructuring Multilayer Perceptrons based on Information-Theoretic Modeling of a *Partido-Alto*-based Grammar for Afro-Brazilian Music: Enhanced Generalization and Principles of Parsimony, including an Investigation of Statistical Paradigms

Mehmet Vurkaç
Portland State University

Follow this and additional works at: https://pdxscholar.library.pdx.edu/open_access_etds

Let us know how access to this document benefits you.

Recommended Citation

Vurkaç, Mehmet, "Prestructuring Multilayer Perceptrons based on Information-Theoretic Modeling of a *Partido-Alto*-based Grammar for Afro-Brazilian Music: Enhanced Generalization and Principles of Parsimony, including an Investigation of Statistical Paradigms" (2011). *Dissertations and Theses*. Paper 384.

<https://doi.org/10.15760/etd.384>

This Dissertation is brought to you for free and open access. It has been accepted for inclusion in Dissertations and Theses by an authorized administrator of PDXScholar. Please contact us if we can make this document more accessible: pdxscholar@pdx.edu.

Prestructuring Multilayer Perceptrons based on Information-Theoretic Modeling of a
Partido-Alto-based Grammar for Afro-Brazilian Music: Enhanced Generalization and
Principles of Parsimony, including an Investigation of Statistical Paradigms

by

Mehmet Vurkaç

A dissertation submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy
in
Electrical and Computer Engineering

Dissertation Committee:
George G. Lendaris, Chair
Douglas V. Hall
Dan Hammerstrom
Marek Perkowski
Brad Hansen

Portland State University
©2011

ABSTRACT

The present study shows that prestructuring based on domain knowledge leads to statistically significant generalization-performance improvement in artificial neural networks (NNs) of the multilayer perceptron (MLP) type, specifically in the case of a noisy real-world problem with numerous interacting variables.

The prestructuring of MLPs based on knowledge of the structure of a problem domain has previously been shown to improve generalization performance. However, the problem domains for those demonstrations suffered from significant shortcomings: 1) They were purely logical problems, and 2) they contained small numbers of variables in comparison to most data-mining applications today. Two implications of the former were a) the underlying structure of the problem was completely known to the network designer by virtue of having been conceived for the problem at hand, and b) noise was not a significant concern in contrast with real-world conditions. As for the size of the problem, neither computational resources nor mathematical modeling techniques were advanced enough to handle complex relationships among more than a few variables until recently, so such problems were left out of the mainstream of prestructuring investigations.

In the present work, domain knowledge is built into the solution through Reconstructability Analysis, a form of information-theoretic modeling, which is used to identify mathematical models that can be transformed into a graphic representation of the problem domain's underlying structure. Employing the latter as a pattern allows the researcher to prestructure the MLP, for instance, by disallowing certain connections in

the network. Prestructuring reduces the set of all possible maps (SAPM) that are realizable by the NN. The reduced SAPM—according to the Lendaris–Stanley conjecture, conditional probability, and Occam’s razor—enables better generalization performance than with a fully connected MLP that has learned the same I/O mapping to the same extent.

In addition to showing statistically significant improvement over the generalization performance of fully connected networks, the prestructured networks in the present study also compared favorably to both the performance of qualified human agents and the generalization rates in classification through Reconstructability Analysis alone, which serves as the alternative algorithm for comparison.

Dedicated to my mother, Sabiha Tuğcu Vurkaç

ACKNOWLEDGMENTS

I am grateful to many people for their help while I worked on this dissertation. First, I would like to thank my mother, Sabiha Tuğcu Vurkaç, for more than twenty years of emotional and financial sacrifice and support as I pursued my higher education away from home, and for a lifetime of love, friendship, teaching, guidance, and music.

This dissertation would not have been possible without the insights, encouragement, mentoring, wisdom, inspiration, confidence and optimism provided by (and the support and open-mindedness of) my adviser, teacher, mentor, and ally Dr. George G. Lendaris. Dr. Lendaris has shown a level of care and engagement in both my general academic, scientific, and intellectual growth, and specifically my research and dissertation well beyond what anyone would expect. He is a true teacher and mentor.

This dissertation also would not have been possible without generous and timely gifts of housing and computing resources from George Karagatchliev and Lisa Brandt Heckman (respectively) that made my continued work possible at the most difficult times of my student career. I'm similarly indebted to Eric Egalite and Joe Reid of OIT for automation and analysis assistance, respectively, and the valuable insights gathered through years of discussions and music-making with Derek Reith. I am also grateful to Woods Stricklin, Patryk Lech, Travis Henderson, Anita Rodgers, Hank Failing, Gary Beaver, Michelle Thayer, Cory Troup, Dr. Bruce Barnes, Andrew Zvibleman, Dr. Xin “Ryan” Wang, and my aunt Semra Bastıyalı for critical assistance at crucial points in the PhD process.

In addition, I would like to specially mention the vital role played by Christina Luther, friend and SEVIS adviser, in keeping me legal and on track.

Also as a primary contributor, I'd like to thank the emergent complexity that resulted from the billions of years of evolution that enabled my wet neural network to engage in such a pursuit as this dissertation.

The presentation of this dissertation and associated work was greatly improved thanks to input from Woods Stricklin, Sabiha Vurkaç, Dr. Eve Klopff, Dan Craver, Dr. Ahmet Müfit Ferman, and the PSU Writing Center, and due to the technical assistance of Dr. Serap Emil (literature review), Tamara Turner and İhsan Tunç Çakır (notation), and Buğra Giritlioğlu (music computing).

I would also like to thank Dr. Melanie Mitchell for inspiring me to go into Computational Intelligence; Brian Davis, Andrew Hartzell and Derek Reith of *Lions of Batucada* for providing the environment for samba to become an integral part of my life; and once again, Dr. Lendaris for bringing the two realms together in my research.

I would also like to thank Dr. Brad Hansen (for his all-around knowledge of the topics, and the efficiency and legitimacy he has brought to my work), Dr. Douglas V. Hall (for being a true teacher and role model), Dr. Dan Hammerstrom (for guidance at various stages, and for introducing me to Jeff Hawkins' work), Dr. Marek Perkowski (for his excitement, encouragement, and many research sources), Dr. Fu Li (for peppering his lectures with invaluable insights), Mestre Jorge Alabê (for all his encouragement, musicality, private lessons, workshops, brain dumps, and *repique* solos),

“long-lost brother” Mark Lamson (for the fourth clave-direction category and for verifying my conceptualization of rhythm), Michael Spiro (for asking the tough questions, and for the IMD factor), Dr. Rob Daasch (for directing me to Dr. Lendaris), Dr. David Glenn of Whitman College (for believing in and supporting me, and giving me time and resources with which to learn to play), Dean/Professor/Cellist Robert Sylvester (for helping me recognize the breadth of my musical experience), Dr. James Morris (for recognizing my teaching ability, which in turn helped finance my studies and prepared me for a teaching career), Dr. Malgorzata Chrzanowska-Jeske (for support and guidance), Professor and Chair Harold Gray (for being a resource and friend), Dr. Martha Balshem (for helping me begin to learn how to conduct research in Ethnomusicology), Derek Reith (first *tamborim* teacher), Brian Davis (first *pandeiro* teacher), Andrew Hartzell (first *surdo* and *repique* teacher), Chris Perry (first reviewer of my clave tutorial), Tobias Manthey (for clave vigilance), Andy Sterling, David Huerta & Jesse Brooke (for clues to 12/8 directionality), Dr. Michael Cummings & Dr. Bill Becker (for early thesis-writing advice), İhsan Tunç Çakır (for the proper way to write standard European music notation), Emily Brown, Steve White and Matthew Stanbro (for “getting it”), Sue Firpo (for substantial assistance with the paperwork and logistics of financing the early years of my PhD), and the following for their work on the golden-ear and benchmark sessions: Tobi Lehman, Ron Scroggin, Lisa Brandt Heckman, Krasi Nikolov, Chaz Mortimer, Tobias Manthey, Rafael Otto, Jake Pegg, John Jenness, Tofer Towe, Gary Beaver, Rachel Sandy, Cory Troup, Michelle Becka and Renata Secco.

For the information-theoretic aspect of this work, I would like to thank Dr. Martin Zwick and Joe Fusion, along with Better World roadside service, and Duane at Farwell Towing.

Furthermore, since this dissertation is truly my life's integrated work, I would like to express my gratitude to Professors David Weber, Don Moor, and Peter Nicholls (for changing my brain for the better); Michelle Thayer; Cingöz Canavar Charlie Monster Vurkaç (my sole companion throughout my candidacy); Torimaru Yumi Octopus (for the subtleties of *dentou geinou* and examples of Japanese notation systems), Cheryl Ramette (for taking the step that made progress possible when everything was about to fall apart), Dr. Candyce Reynolds, Michelle Gano (for helping me through the financial bureaucracy of PSU), students of my Saturday samba class (for all their questions, enthusiasm, and friendship), Danny Norton (clave-conscious Drum Buddy virtuoso), Devon Halloran and Emily Philips (for emergency academic/clerical assistance), The PSU University Studies 2007–2008 Mentor Community (especially Prof. Teresa Taylor, Kim Heidenreich, Nancy D'Inzillo, Jason Baidenmann, Nathania Cha, Christina Overturf, Chris Howell, Cara Glennon, Christie Toth, and Kate Gentry), session musician and childhood friend Yavuz Çetin (for opening the door to *so much* music), İzi Eli (first music teacher), Faik Tuğcu (my grandfather and first music mentor), my aunt Nemika Tuğcu (role-model for strength and positivity), my aunt Solmaz (Vurkaç) Başpınar, Dr. James McNames (for scientific vigilance); Nikki and Shiva Thayer; Lions of Batucada, Mais Que Samba, California Brazil Camp, toyboat toyboat toyboat, Obsessive Percussive Disorder, Whatsoever Creepeth, Kafana Klub,

Takohachi, Mirah Yom Tov Zeitlyn and Emily Kingan; multi-talented drummer/composer/role-models Roger Taylor, Memo Acevedo, Airtó Moreira, Stevie Wonder, Phil Collins, Vanderlei Pereira, Mauricio Zottarelli, and Yoshiki Hayashi; Igor Stravinsky, Anton Webern, Karlheinz Stockhausen and John Cage (for advancing tonal and timbral democracy); Scott Wardinsky, Engin Gürkey and Jozi Levi; Paul R. Bergmann, Tony Williams and Dick Schalk; Okay Temiz, Trilok Gurtu, Giovanni Hidalgo, Doudou N'Diaye Rose, and Gregg Bendian (five inspiring drummers who took the time to talk to me and encourage me); Robert College (my middle and high school in Turkey, for instilling American-style skepticism), Whitman College, Portland State University (for giving me the chance many other institutions may not have), Oregon Institute of Technology (Anna Bailey, Professors Jamie Zipay, Paul Dingman, Valerie Ball, and Matt Gildehaus, Casey Coulson, Dr. Bruce Barnes, Dr. Claudia Torres-Garibay, Dr. Xin “Ryan” Wang, and Dr. Mateo Aboy, who supported me and my work in tangible ways while I completed the research for this dissertation); the FSM and the IPU; and most importantly, Mustafa Kemal Atatürk (for creating modern, secular Republic of Turkey, making my-life-as-I-know-it *possible*).

In closing this rather long acknowledgement, I'd like to also thank the people who shaped my musical *or* academic life the most, respectively: John Deacon, Roger Taylor, Freddie Mercury and Dr. Brian May (of Queen), and Yılmaz Vurkaç (my late father, for the drum sticks, the love of books, obsessiveness, and who knows what else).

TABLE OF CONTENTS

ABSTRACT.....	i
DEDICATION	iii
ACKNOWLEDGMENTS.....	iv
LIST OF TABLES	xvii
LIST OF FIGURES.....	xviii
PREFACE	xxii
CHAPTER I. INTRODUCTION	1
1.1 Problem Statement.....	1
1.2 Motivation and Purpose.....	2
1.3 Objective of the Study.....	12
1.4 Research Conducted	13
1.5 Contributions	15
1.6 Significance and Rationale	16
1.7 Assumptions and Hypotheses	18
1.8 Scope, Limitations, and Constraints	19
1.9 Teacher Models	20
1.10 Selection of Holdout Data.....	21
1.11 Output Encoding.....	22
1.12 Number of Hidden Layers.....	23
1.13 Number of Hidden-Layer Elements	23
1.14 Criteria for Model Evaluation and Selection in RA	25
1.15 Output Thresholding in MLPs.....	39
1.16 RA-Search Heuristics.....	43
1.17 Definitions.....	47
1.17.1 Multilayer Perceptron (MLP)	47
1.17.2 Generalization	48
1.17.3 Clave Direction	49
1.17.4 Tactus and Tatum	50
1.17.5 Syncopation and Offbeatness	52

1.17.6 Attack-Point Rhythm	55
CHAPTER II. BACKGROUND	56
2.1 Science and Computational Intelligence	57
2.2 An Overview of Machine Learning and Statistical Learning	62
2.2.1 Types of Cross-Validation	66
2.2.2 Holdout Cross-Validation: The Basics	69
2.2.3 Holdout Cross-Validation for Design	69
2.2.4 k -Fold Cross-Validation: The Basics	69
2.2.5 k -Fold Cross-Validation for Design	70
2.2.6 Statistical Soundness	70
2.2.7 Holdout Cross-Validation for Assessment	71
2.2.8 k -Fold Cross-Validation for Assessment	71
2.2.9 Motivation for Cross-Validation	71
2.2.10 The Overall Holdout	72
2.2.11 Performance Metric	72
2.3 Artificial Neural Networks	76
2.3.1 Historical Background on Neural Nets	76
2.3.2 The General Algorithm for the Delta Rule	83
2.3.3 A Brief Survey of Other Neural-Net Types	91
2.4 Reconstructability Analysis	101
2.4.1 RA and <i>Occam3</i>	103
2.4.2 Historical Reference	116
2.4.3 RA Formalism	116
2.4.4 Aspects of Statistics Relevant to RA	118
2.5 Clave and Clave Direction, with a note on labels	122
2.6 Information Theory, Reconstructability Analysis, Multilayer Perceptrons, and the Clave Concept: The Connection	130
2.7 Researcher's Background and Qualifications	132
2.7.1 Academic Background as Student	132
2.7.2 Academic Background as Instructor	133

CHAPTER III. LITERATURE REVIEW	134
3.1 Neural-Network Prestructuring	135
3.2 Literature Review for Neural-Network Prestructuring	136
3.3 Current Practice in Neural-Network Optimization	140
CHAPTER IV: METHODOLOGY	150
4.1 Introduction	150
4.2 Notation and Data	151
4.2.1 <i>Occam3</i> Notation	151
4.2.2 Music Notation	152
4.3 Nature of the Data	153
4.4 The Determination of Data Splits	161
4.5 Experiment-Design Rationale	162
4.5.1 Factorial Designs	164
4.5.2 Statistical Power, Significance and Hypothesis Testing	165
4.5.3 Control Design	175
4.5.4 Assumptions, Justified and Unjustified	176
4.6 Experimental Process	177
4.6.1 The General Experimental Process	177
4.6.2 The Processes of Data Acquisition and Preparation	180
4.6.3 Data Acquisition: Classification	183
4.6.4 Training-and-Test Regimes	195
4.6.5 RA Search, Fitting and Selection	199
CHAPTER V. APPLICATION AREAS	210
5.1 Neural-Network Prestructuring	210
5.2 Music Training	212
5.2.1 The Growing Relevance of Clave	212
5.2.2 A Note on Terminology	214
5.2.3 Introduction to Clave	217
5.2.4 Systems of Musical Harmony	217

5.2.5 Clave as a Principle of <i>Rhythmic</i> Harmony.....	218
5.2.6 A Closer Look at Some Intricacies of Clave.....	220
5.2.7 Why is the Study of Clave Important?.....	224
5.3. Clave and Music Technology.....	225
5.3.1 Why a High-Tech Solution to Clave Training?	225
5.3.2 Current Research and Appropriate Technologies for Machine Recognition of Clave Direction.....	227
5.3.3 A high-level framework for realizing automated recognition of clave direction.....	229
5.4. Clave Analysis as a Music-Production Feature	232
5.4.1 Music-Making.....	232
5.4.2 Music Training.....	233
5.4.3 Music Teaching	234
5.5 Conclusion.....	235
5.6 JMTE-Suggested Citation for the Published Article.....	236
CHAPTER VI: FINDINGS/RESULTS.....	237
6.1 Results of Neural-Net Design.....	237
6.2 Tables of Numerical Results.....	238
6.3 Preliminary Statistical Analysis	246
6.4 Full Statistical Analysis	255
6.5 Discussion of Findings.....	265
CHAPTER VII: CONCLUSION.....	271
7.1 Prestructuring Performance.....	271
7.2 Statistical Significance	272
7.3 Implications for the Use of Reconstructability Analysis in the NN Field.....	273
7.4 General Implications for Neural Networks	274
7.5 Musical and Music-Technological Insight.....	274
7.6 Recommendations Concerning the Methodology of Future Studies and Focus of Future Development.....	275
7.7 Limitations.....	277
7.8 Possible Sources of Error	278

7.9 Shortcomings of the Dissertation	279
7.10 Strengths of the Dissertation.....	280
7.11 Summary.....	281
REFERENCES	283
LIST OF APPENDICES.....	296
APPENDIX A: A Cross-Cultural Grammar for Temporal Harmony in Afro-Latin Musics: Clave, <i>Partido-Alto</i> , and Other Timelines.....	297
A.1 Introduction	297
A.2 The Basic Claves (Clave-Propor).....	302
A.2.1 The (Afro-Cuban) <i>Son</i> Clave: A seemingly obvious starting point.....	302
A.2.2 The So-Called Brazilian (<i>Bossa</i>) Clave	304
A.2.3 A Graphical Analogy for Clave Direction.....	306
A.2.4 The Clave Prototypes: Fundamental Rhythms for Clave Direction.....	308
A.2.5 Wide-Sense Clave Consciousness.....	311
A.2.6 Relative Nature of Clave Direction in Other Afro-American Rhythms.....	315
A.2.7 Conclusion	321
References for Appendix A.....	323
APPENDIX B: Information Theory.....	326
B.1 Uncertainty and Information.....	326
B.2 Entropy.....	331
B.3 Transmission, Mutual Information, Information Distance and the Kullback- Leibler Divergence	333
References for Appendix B.....	338
APPENDIX C: Literature Review of Music Information Retrieval and Digital (Audio) Signal Processing.....	339
C.1 The Reviews.....	340
C.2 Conclusion	369
References for Appendix C.....	371
APPENDIX D: Statistics, Social Responsibility, and the Enhanced Scientific Method	374
D.1 Why Statistics?	374

D.2	The Role of Statistics in the Scientific Method: Statistics as Meta-Science.....	375
D.2.1	Misuse of Statistics in Science, Medicine and Technology.....	381
D.2.2	Misuse of Statistical Techniques in Model Evaluation.....	386
D.2.3	What Can Be Done?.....	387
D.3	Statistics, Critical Thinking, Science, and Social Responsibility	388
D.4	The Scientific Method, Past and Present.....	403
D.4.1	Accuracy.....	406
D.4.2	Objectivity	406
D.4.3	Skepticism.....	407
D.4.4	Open-mindedness.....	407
D.4.5	Additional Concerns, Techniques, and Principles	407
	References for Appendix D	410
	APPENDIX E: The Data-Acquisition Process	416
E.1	Data Acquisition: Golden Ears	420
E.2	Data Acquisition: Classification	423
E.3	Data Acquisition: Golden Ears	428
E.4	Bootstrap Process: Misclassification and NN Feedback.....	430
E.5	Bootstrap Process: Context-Dependent Patterns.....	443
E.6	Final Approach	443
	References for Appendix E.....	444
	APPENDIX F: Brazilian Glossary, Clave Terminology, Musical Definitions, and Samba Concepts.....	445
F.1	Author's Original Samba Glossary.....	445
A.....		445
B.....		446
C.....		447
D.....		450
E.....		450
F.....		450
G.....		450

L	451
M	451
O	452
P	452
Q	454
R	454
S	455
T	457
V	458
X	459
Z	459
F.2 Clave Terminology throughout the Diaspora	459
F. 3 Musicological Definitions, Brazilian, European and Global	459
F.3.1 Definitions of Music	460
F.3.2 The Elements of Music	461
F.3.3 Time, Rhythm and Meter	464
F.3.4 Brazilian Musical Definitions from Jorge Alabê	465
References for Appendix F	467
APPENDIX G: Northern/Non-Northern vs. Western/Non-Western	469
APPENDIX H: Occurrences of Clave-Type Patterns in Music around the World, and Recommended Listening	471
H.1 Examples of Clave Direction in Brazilian Musics	471
H.2 Accessible Examples of Clave in other African-based Music, and Jazz	473
H.3 Clave-type Timelines or Patterns found in Non-Clave Musics	475
H.4 Recommended Listening in African and Afro-Latin Music	480
APPENDIX I: Author's Musical Background and Qualifications	482
APPENDIX J: A Survey of Notational Conventions and Preferences for Meter and Time Signature for Musics of the African Diaspora	483
J.1 Purpose	483
J.2 Data	483

J.3 Findings	486
J.4 Conclusion	495
References for Appendix J	496
APPENDIX K: The <i>automate.dat</i> Script for Neural-Net Experiments in NeuralWare's NeuralWorks	499
APPENDIX L: The <i>AutomateNeuralWorks.vbs</i> Script for Exercising Network Configurations with Multiple Seeds	500
APPENDIX M: C Code for the Top-Generalizing BIC-based Network.....	506
APPENDIX N: Sample Data in the Firm-Teacher Context	511
APPENDIX O: Random Seeds	512

LIST OF TABLES

Table 1: Types and Applications of Cross-Validation.....	67
Table 2: Data Breakdown for Null-Hypothesis Example.....	113
Table 3: Single determining factor.....	113
Table 4: Single determining factor; reversed correlation.....	114
Table 5: Output Dependence on Both Factors.....	115
Table 6: Breakdown of the full data set into clave-direction categories and gross membership degrees.....	159
Table 7: Combinations of RA Search Strategies and Criteria.....	200
Table 8: Initial Results.....	241
Table 9: A selection of Generalizing Ratio values for the five “finalist” neural-net types as exercised by different random-number seeds.	242
Table 10: Generalizing Ratios used in the initial ANOVA, with arbitrary identifiers. ..	247
Table 11: One-way ANOVA: Ratio versus Method.....	248
Table 12: Grouping Information Using the Dunnett Method of ANOVA	250
Table 13: Confidence Intervals for Treatment Mean minus Control Mean	250
Table 14: 95% Bonferroni Confidence Intervals for Standard Deviations.....	254
Table 15: 95% Bonferroni Confidence Intervals for Standard Deviations for the Full Analysis.....	262
Table 16: Welch's ANOVA, as compared with Dunnett ANOVA.....	264

LIST OF FIGURES

Figure 1: High-level block diagram of overall research flow, showing the design data, the structures compared for their generalization performance, and the primary comparison technique (ANOVA). Note: There are more than two RA-based prestructured neural networks, as explained in the section on RA search heuristics (Section 1.16), but only two were “finalists” based on their performance in the preliminary experiments. Also, the shading is an artifact of Visio, and not meaningful.	17
Figure 2: Overfitting in terms of model complexity, an experimental reproduction of the classic plot.	31
Figure 3: Change in degrees of freedom versus model "level." Level, also related to model complexity and search time, is the number of input variables in any single-component model, where a component is a collection of variables related in terms of their joint effect on the output.	32
Figure 4: Overfitting and cumulative significance level	33
Figure 5: Overfitting and change in model Entropy	34
Figure 6: Change in AIC, change in BIC, and overfitting: A straight line has been added where $\Delta AIC = \Delta BIC = 0$, showing that BIC may indicate the beginning of the training/test performance split, but that AIC clearly marks the drop in test performance.	35
Figure 7: AIC, BIC, percent change in Entropy (Uncertainty), and overfitting in an all-model search.	37
Figure 8: Entropy, Information, significance level, and overfitting in an all-model search	38
Figure 9: Belizean “3-2” clave (top) and the common <i>son</i> clave (bottom) in the same “direction.”	50
Figure 10: Two parts (to be played simultaneously) for demonstrating tatum-finding.	51
Figure 11: Even though there are no 16th-note inter-onset intervals (IOIs, [52, p. 4]) on either instrument, the tatum for the overall piece is of 16th-note duration because of the interaction between the two parts in the second bar.	51
Figure 12: A few of the attack-point patterns presented to participants in the human-benchmark studies; the shading and large font size are for participants’ ease of reading. Support for this representation is found in industry [60], academia [11], and among professional musicians [61].	55
Figure 13: Cohen's diagram of scientific progress (recreated).	59
Figure 14: A possible partial classification of engineering approaches to search, optimization, prediction (forecasting) and classification (pattern recognition) problems. These divisions, even between Machine Learning and Statistical Learning, are not	

absolute—there is overlap among many of these categories, both in terms of methods and in terms of the underlying philosophies.....	63
Figure 15: A partial family tree of subfields in Computational Intelligence. Not shown due to space constraints are the Generalized Regression Neural Network (GRNN), Counterprop, CMAC, the Infomax network, various types of Recurrent Neural Networks, and some Swarm algorithms.....	68
Figure 16: The Perceptron.....	81
Figure 17: Changing error surfaces (shown in two dimensions) for stochastic approximation to gradient descent [82].....	85
Figure 18: Single-Layer Neural Network.....	86
Figure 19: Basic Bidirectional Associative Memory (BAM) setup.....	98
Figure 20: A 2 x 2 x 2 contingency table shown in shorthand notation (above) and explicit probability notation (below).....	108
Figure 21: Example of calculated two-variable model values based on the marginals of the observed distribution (data).....	109
Figure 22: The Lattice of Structures for (only) four variables, taken from [47].	117
Figure 23: The set of all possible mappings (SAPM, light grey oval), the performance subset (PS, dark grey oval), and the desired mapping (DM, dark dot) for two hypothetical neural networks	139
Figure 24: Modular (a), and fully connected (b, c) neural networks.....	140
Figure 25: 16 th -note subdivisions and the corresponding <i>Occam3</i> variable names.	152
Figure 26: The settings for a narrow <i>Occam3</i> RA search (MBR–a1) of mathematical models with statistical model evaluation and selection, to determine the number of variables (which correspond to “levels” in a loopless search) for the subsequent step, along with provisions (in the data set and the search settings) for RA generalization results to be used in the ultimate comparison that is the goal of the present research.	201
Figure 27: The “rotate” step (MBR–a2) of the bore–rotate search strategy, with search width (the number of models to be considered at each level) increased to 20.....	201
Figure 28: All-model BIC (not AIC) search (MBR–a3) with high width and depth for feature-reduced inputs to determine the best statistical model of the “design-all” data set.	202
Figure 29: Narrow loopless-model search with Information as the criterion (MBR–i1) to determine the number of input variables (loopless levels) to be included in the subsequent “rotate” search.....	203
Figure 30: Search MBR–i3 for “design-all” data that resulted in a “large” model.....	205
Figure 31: ABE–b1 search settings for the all-model bore–expand BIC search.....	206
Figure 32: ABE–b2, the “expand” portion of the BIC all-model bore–expand search. The result of this search is the third candidate model.....	207

Figure 33: ABE-i1 set at 200-deep after the 500-deep search crashed with no results after several days.	208
Figure 34: Prestructured neural network (multilayer perceptron) based on the BIC-based model resulting from the all-model boreexpand (ABE) search.	211
Figure 35: <i>Partido-alto</i> , the rhythmic basis of samba, in 3-2 clave direction.	216
Figure 36: The son clave: the Cuban name for a common pattern in Brazilian, Caribbean and West African music, as well as early rock'n'roll and jazz. It is the best-known of all clave patterns.	217
Figure 37: A possible samba <i>surdo</i> pattern in 2-3 clave direction: Comparing numbers of onsets in each half is evidently not helpful.	220
Figure 38: One of the standard <i>surdo</i> patterns in Rio-style samba employed by G.R.E.S. Mocidade Independente de Padre Miguel and many other samba schools. (The notation disregards samba swing.) This pattern matches the 3-2 <i>bossa</i> clave or is “in 3-2.” Note that there are two onsets in the first half and three in the second. This contradicts the simple-minded approach of counting onsets in clave halves, a practice that possibly owes its popularity to the ubiquitousness of the <i>son</i> clave. On the other hand, reconciling this with the idea that this pattern is in 3-2 clave orientation is straightforward and free from contradiction once one becomes familiar with the relative nature of clave offbeatness (Appendix A, submitted).	221
Figure 39: A common accent pattern that can be heard in both traditional and modern Afro-Brazilian music, this pattern is (perhaps surprisingly) typically associated with the 3-2 clave direction even though the second half of the pattern is identical to the 3-side of both <i>son</i> and <i>bossa</i> <i>clave</i>	221
Figure 40: A contrived (not traditional, but still likely) pattern of accents in 2-3 clave orientation: Once again, note that the first half has three onsets while the second half has two. Yet, the overall pattern is in 2-3, not 3-2. This is similar to the point made in Figure 9 above and Figure Twelve in Appendix A.	221
Figure 41: A typical 3-2 agogô (bell) pattern in Rio-style samba: Yet again, counting the onsets (even if one uses the foreshadowing concept) is of no help in determining clave orientation.	222
Figure 42: The 3-2 agogô pattern of Figure 41 becomes 2-3 when phase-shifted by one onset.	222
Figure 43: The accent pattern (ignoring grace notes) for the caixas (snare drums) in the Rio samba school G.R.E.S. Estácio De Sa [167, p. 51]. Since the only note breaking the steady pattern of eighth notes (onbeats) is the last one, that is the only clave-indicating onset, and the pattern is in keeping with 3-2 samba direction.	222
Figure 44: One possible notational abstraction of the standard caixa pattern in the Mangueira samba school: This pattern matches the 3-2 patterns in the rest of the ensemble. The only clue as to the sense of clave direction is in the difference between	

the offbeat sections following the second and fourth downbeats ([2]-E-AND vs. [4]-E-A). This is how clave is subtle: A single 16th-note shift, i.e., the placement of one onset (out of ten), is enough to set the clave direction.....	223
Figure 45: The second bar of a two-bar phrase used by the third <i>surdo</i> in the Rio samba school Império Serrano [167, p. 53], this is a good indication of the subtlety and relativity of samba offbeatness. A method of cancellation of like onsets in corresponding portions (similar to cancelling off terms in algebra) yields the pattern of Figure 46, which is clearly in 3-2, while the original pattern shown here initially appears to establish no clave direction.....	223
Figure 46: The pattern of Figure 45 after cancellation, which results in a pattern closely related to the standard 3-2 <i>surdo</i> pattern in Figure 38.....	223
Figure 47: The audio waveform for one bar of 3-2 son clave, as captured in MATLAB.	227
Figure 48: A system-level data-processing stream (with human interaction) for identifying clave direction in MIDI data received in real time.....	230
Figure 49: A system-level data-processing stream (with or without human interaction) for identifying clave direction in multi-channel recorded or multi-instrument live audio.	232
Figure 50: 95% Confidence Intervals for One-Way ANOVA, individual comparisons of Prestructured Network Types against Fully Connected.	249
Figure 51: Confidence Intervals for Treatment Mean minus Control Mean.....	251
Figure 52: Residual Plots for Ratio.....	252
Figure 53: Test of Normality of Distribution.....	253
Figure 54: 95% Bonferroni Confidence Intervals for Standard Deviations.	254
Figure 55: Ratio versus Method for Full Analysis.....	255
Figure 56: Means, Standard Deviations, and 95% Confidence Intervals for the Full Analysis.....	256
Figure 57: Dunnett ANOVA showing that all four network types are statistically significantly different from fully connected.....	257
Figure 58: Bonferroni Adjustment showing that all four network types are statistically significantly different from fully connected.....	257
Figure 59: Dunnett 95% Confidence Intervals and <i>P</i> Values.	258
Figure 60: Bonferroni 95% Confidence Intervals and <i>P</i> Values.....	258
Figure 61: Tukey pairwise comparisons show statistically significant differences among all means. (All Tukey <i>P</i> values are arbitrarily close to 0.).....	259
Figure 62: Results of the tests of normality for the full analysis.....	261
Figure 63: The four outliers (out of 1935 data points) that lead to doubts on accepting the normality assumption for ANOVA.	262
Figure 64: Comparison of Variances for the Full Analysis.....	263

PREFACE

This dissertation contains material that is not typically encountered in Electrical & Computer Engineering research. The atypical material includes an inquiry into (and claims about) Science, Statistics, and their connection to critical thinking and social responsibility.

The reason I chose such a broad approach (asking questions about Probability, Statistics, inference, confirmation, and Science itself, and pursuing these questions down the rabbit holes of Evolution, History and Psychology) has to do with four key factors: the influences of two current members and one former member of my dissertation committee, and that of my teaching job at the time of much of my proposal research.

When I first approached Graduate Studies representative Dr. Hansen about being on my committee, one of the things he said was that I should have a testable hypothesis. At that time, I was co-developing and co-teaching the “Knowledge, Rationality and Understanding” (KRU) course in Portland State University’s general-education program, University Studies (UnSt). Even though I had a BA in Physics, it was this course that introduced me to the idea of *what Science is*, and of all the things that it isn't. I was delighted at the power that mathematical and logical rigor (in the form of statistical techniques, Bayes theorem, and so on) could bring to an investigation of facts, regardless of whether those facts were natural, societal, or technological.

At the same time, I was horrified at the awful things being done in the name of science, or more typically, in the name of *doing good*. Some of these things fall under the headings of pseudo-science and bad (poor) science. Pseudo-science, of course, is not

Science at all. Just as harmful is when real science is not practiced well (due to carelessness, laziness, time constraints, lack of funding, or just methodological deficiency), the results obtained can have damaging effects on society in very real ways (such as in terms of health and public safety), if not debunked through independent verification. My broadly inquisitive approach to Neural Networks as well as the proposed grammar of *samba carioca* is due to this shared responsibility that scientists (which ideally include all PhDs) have toward society and the pursuit of truth.

Thirdly, in teaching KRU, I learned about unresolved fundamental issues in *good* science, from publication bias to the problem of induction, to whether there really is a scientific method in practice, to Occam's Razor, to the conflicting interpretations of probability, to the all-too-common misinterpretations of statistical significance, one of which I was committing just as I taught my students to avoid another.

Between teaching KRU seven times in three years and scrutinizing the textbooks on statistical techniques Dr. McNamara lent to me, Dr. Hansen's warning struck me hard. I felt that *I must do good science*, nothing less. (Perhaps more, but never less.)

The Systems Science program, headed by my dissertation adviser, Dr. Lendaris provided another incentive, and indeed a requirement, towards approaching my research in a systems-oriented and systems-wide manner. So did the encouragement and appreciation of my comprehensive-exam-committee member Dr. Mitchell when I proposed a preliminary research plan that included readings in the developmental biology, psychology and physics of music).

As a result, my research proposal (as accepted) included responses to

- questions posed,
- challenges presented,
- requirements made, and
- encouragements given

over the years by (in chronological order)

- Dr. Jenq (my first adviser in Electrical & Computer Engineering who had specific concerns over the range of applications of my proposed research),
- Dr. Mitchell,
- Dr. Lendaris (systems perspective),
- Dr. Hansen (cultural scope and scientific validity; close-miss idea),
- Dr. McNames (statistical methods and assumptions)
- Dr. Hammerstrom (consideration of time-dependent approaches like HMMs and HTM)
- Dr. Hall (consistent support throughout my MS and PhD),
- Dr. Zwick (Reconstructability Analysis and the “lousy-control” idea)
- Dr. Perkowski (the enthusiasm he brought to my presentations and our discussions).

As a result of this broad basis of support and guidance, I studied, to the best of my ability and resources, relevant material in the fields of

- Statistics,
- Information Theory,
- Philosophy of Science, Logic and Epistemology (as required for any inquiry into Statistics),
- Neuroscience and Cognitive Science (essential background to any study of Artificial Neural Networks),
- Cognitive, Experimental and Evolutionary Psychology (intimately connected to Neuroscience),
- Developmental and Evolutionary Biology,
- Psychophysics (superset of Psychoacoustics, and overlapping with Music Theory),
- Musicology (theories of form and structure, music education, and especially rhythm and meter),
- Sociology of Music (primarily to justify my desired research area to early committee members and also as a result of my exposure while co-teaching Popular Culture in University Studies),

- Area Studies (the history, economics, politics and culture of Africa and Latin America, due to the connections between Afro-Brazilian music, other African Diasporan music, and the transatlantic slave trade),
- Ethnomusicology (to the extent necessary to understand the issues involved in studying the music of a culture other than my own so as to develop sufficient understanding to draw my data from that realm),
- Mathematics (the geometric and computational basis for Computational Ethnomusicology, as exemplified by the works of Toussaint and collaborators)
- Computer Science (scripting and programming for neural nets and evolutionary algorithms),
- Computer Engineering (Statistical Learning, DSP and Statistical Signal Processing),
- Soft Computing (Fuzzy Logic),
- Systems Science (Reconstructability Analysis, Complexity, Emergence, Systemness, and the concept of wholes and parts), and
- Music Technology, musical applications of Robotics, and Information Retrieval (the context for the practical application that may arise from the music aspect of my research).

My research focus and the bulk of my effort changed (from 2001 to 2011) from the Peeling Algorithm for de-reverberation, to clave in Afro-Latin Music, to Neural Networks, to Information Theory and mathematical modeling, and finally to statistical techniques.

In the course of my growth as a scientist, I recognized seven levels of statistical comprehension. These are represented below with characteristic statements that a person at that level of sophistication may say:

1. I see stats, so I believe what they say.
2. Statistics are all lies and manipulation.
3. When I see stats, I ask questions about sample size, confidence intervals, assumptions and design (randomization, blocking, etc.) before taking them seriously.

4. Even if good research design was employed, statistical significance must be investigated before I pay any attention.
5. Even if good research design was employed, statistical significance or confidence intervals or Bayesian inference or support for assumptions is needed to take the results seriously.
6. By looking at the experimental design of a study, conditions and assumptions, I can tell the difference between meaningful design in terms of statistical-power/statistical-significance and a weak, faulty design, as well as the ways in which power and significance techniques should be applied differently in the social sciences, physical sciences and technological fields.
7. In addition to being able to do [6], I'm also aware of the phenomenon of shrinking effect sizes, Kuhnian, Popperian, Humeian, and other challenges to scientific method(s), and other meta-level issues in Science and Statistics. I am, therefore, skeptical of anything short of widespread independent verification under a variety of circumstances, whether the question regards antidepressants, acupuncture, M-theory, or neural networks.

These descriptors may not reflect the process others go through, and may show up in a different order or not at all for scholars who are more knowledgeable than me. This is also not the end of the journey of statistical discovery for me. Nonetheless, these are the steps I have either taken so far or observed in people around me.

Thus, you have a dissertation in your hand that is primarily about testing the principle of parsimony in neural-net design, with regard to generalization performance, and as represented by the information-theoretic model-selection process called Reconstructability Analysis.

The data for this investigation comes from a grammar of Afro-Brazilian rhythm for *samba carioca* developed as part of the present research, and currently being proposed to the musicology and Afro-Brazilian-music communities through two articles. One article has been published in *The Journal of Music, Technology & Education* (included in Chapter V). The other article goes into greater detail regarding the proposed grammar of clave direction, and is under review at a journal of musicology.

This proposed grammar is a mapping of inputs (idealized accent patterns) and outputs (rhythm categories). The mapping is exemplified by over 10,000 I/O pairs, compiled from the music literature and instructional books, 19 years of study, and through experiments with master drummers. The 10,000-plus data include a large holdout set used only in the evaluation of the networks' generalization performance.

The experimental design and the analysis methods are based on a rather extensive (though distressingly, but necessarily, incomplete) *survey and investigation of the statistical and engineering literature on model selection, model evaluation, experimental design, and statistical analysis.* These are explained in the body of the dissertation as well as in the appendices. The appendices also include an attempt at justifying the emphasis on statistical design throughout the dissertation. The justification

is given in terms of the obligation of any researcher to uphold the best possible scientific practice, and this obligation is supported by evidence (and well-founded conjecture) based on the evolutionary history and necessity of cognitive biases and logical fallacies.

Cliché as it might be, I hope that the reader enjoys this dissertation—a juxtaposition of technology, science, art and philosophy—a fraction as much as I enjoyed writing it.

Mehmet Vurkaç

October 8, 2011

CHAPTER I. INTRODUCTION

1.1 Problem Statement

The prestructuring of multilayer perceptrons (MLPs) has been shown to be effective in improving their generalization performance under carefully constructed conditions of synthetic data generated by a known logical process.

The goal of the present research is to expand the scope of prestructuring-for-generalization-improvement to real-world problems. Such problems are represented here by a problem domain in which the mapping (the relationship between patterns and their classification) is of a cultural nature. The basis for this mapping (clave direction in its various cultural manifestations) is not fully and clearly known, but the existence of the mapping has been confirmed in the musicological literature (both scholarly and popular) [1–11]. The musical inquiry that forms part of the basis of this study is unique in that the few attempts to define or analyze clave [4, 12] are either too broad in their cultural reach, or focus on analyzing other (mathematical) dimensions of the mapping [11].

Using the aforementioned problem domain, we investigate the generalization performance of prestructured MLPs under the less favorable conditions of a noisy, ill-defined real-world problem.

For purposes of this dissertation, the domain of application will be called *partido-alto*-based clave direction in *samba carioca* (which means “Rio-style samba”). Within this domain of application, we test the applicability of Occam’s Razor (the principle of

parsimony, discussed in Section 1.2) through the Lendaris-Stanley conjecture¹ to the design of multilayer perceptrons in terms of learning and generalization.

1.2 Motivation and Purpose

Characterizing the environment in which neural nets learn is an important thing that many people in the connectionist community usually overlook. They just blithely write a program that produces the environment the network learns in, without ever asking where the *structures* in that environment come from in the first place (Ed Hutchins, cognitive anthropologist, as quoted in [14], p. 174; emphasis in the original).

It is well known, at least in the Computational Intelligence community, that multilayer perceptrons with nonlinear activation functions and a hidden layer can learn any mapping to a satisfactory degree, given enough computing resources. It is also recognized that all the work in AI, CI, ML² and related fields taken together has not come close to getting a glimpse of the emergence of consciousness (although the IEEE has reported that an artificial neural network has been made to show signs of schizophrenia [15]). The emergence of consciousness in machines may require much that is not yet sufficiently understood or implemented, including, but not limited to experience, perception of context, high-level learning, self-directed inquiry, and other aspects related to cognition and awareness.

Structure, as described by Hutchins, is related to one of those elements: context. The work in this dissertation is first and foremost that of prestructuring and its effect

¹ “A conjecture is a proposition that looks true but has never been proved.” [13, p. 11]

² The highly overlapping fields of Artificial Intelligence (AI), Computational Intelligence (CI), and Machine Learning (ML) are listed throughout this dissertation mostly as AI, CI, and ML, respectively.

on generalization, which together are likely to be contributing elements to consciousness. Prestructuring is the two-step procedure of determining the structure of a problem domain and somehow translating that structure to the architecture of the solution (in this case, a neural network). While consciousness is not the goal here, the pursuit of some of its potential components, generalization and context-based structure, has been investigated by means of extracting structure (*Information Theory*) from a real-world problem (*the recognition of clave direction*), and applying the resulting structure to a comparison of prestructured and fully connected neural networks (*Computational Intelligence*). Along the way, it became imperative to understand the pros and cons of different model-selection and model-evaluation technique (*Statistics* and *Philosophy of Science*), both for the prestructuring task and for the final analysis of the networks' performances.

The author's background in the latter two areas is based on co-developing and co-teaching several sections of a critical-thinking course at Portland State University (with three faculty members from the Philosophy Department). The course incorporated elements from formal logic, philosophy of science, cognitive and experimental psychology, probability, and statistics.

The author's interest in and experience with Afro-Brazilian rhythm has a history of about 26 years as listener, 19 years as student, 14 years as performer, 11 years as teacher, and 10 years as researcher.

The author's interest in Computational Intelligence arose initially through curiosity, and then through courses taken at Portland State University with professors

Mitchell, Lendaris and Greenwood (in the departments of Computer Science, Systems Science and Electrical Engineering, respectively) covering the fields of Bayesian Learning, Neural Networks, Evolutionary Algorithms, Fuzzy Logic, and Reinforcement Learning.

The motivation to pursue this research arose out of the convergence of these intellectual, artistic and professional pursuits. ***The intent of the research is primarily to contribute to the field of Computational Intelligence*** through the study of a real-world problem (selected because of the researcher's intimate familiarity with its details and pitfalls). Secondary to that, an additional benefit is to contribute a theory of clave direction to the field of Musicology supported by technological research with strong theoretical foundations, and in the process, to contribute to the technology and practice of music teaching in an area that is not addressed by the established means and channels of music education.

It is a well-known strength of the Neural Networks methodology that an artificial neural network can discover and capture effects that are difficult for conscious human analysis to explicitly describe. This is the reason for using clave direction in research that has the intention of helping improve design practices in the Neural Networks field. At the same time, the use of neural-net technology can serve to expand our understanding of clave direction. These two mutual benefits comprise a sort of bootstrap method where the incomplete function of well-known and well-understood pattern–direction relationships are used to train neural nets, which in turn provide new pattern–direction relationships that augment the experts' understanding of clave (in this

case, as practiced in the Afro-Brazilian traditional idiom). In the end, the final set of I/O relationships constitute a training and a test (holdout) set where high confidence in the relationship of even the most obscure input patterns to output categories allows approaching the problem of neural-net generalization with a very large data set of real-world examples that incorporates, but controls for, noise³ in the data.

Subsequently, information-theoretic modeling is used to aid the process by trading off complexity and generalization to identify models that capture sufficient Constraint to explain observed relationships *and* leave room for generalization without excessive overfitting.

The background in which this dissertation is done, including the design choices made, such as the comparison with human agents and the choice of output encodings for the neural nets studied, reflect the intention of ultimately (post-PhD) developing a musician-training product. This is in addition to the scientific/philosophical and technological intents articulated throughout this document.

³ Noise is of three sources:

- 1) Noise as a result of human error in data entry, which is inevitable,
- 2) Noise due to the nature of the problem domain, *partido-alto* clave direction in *samba carioca*, which has not yet been neatly defined for all contexts in the sense that is widely misunderstood, misrepresented, and only applied in a very small range of possible circumstances by a small number of people-in-the-know whose explanations are vague and sometimes inconsistent, and

- 3) The type of stochasticity described by Hastie et al.:

“Here the function is deterministic, and the randomness enters through the x location of the training points” [29, p. 28].

An example of the first type of noise is the vector 43905. Even in the latest data set arrived at after six passes, 43905 had one bit in one of its output encodings entered wrong (a typographical error).

The second type of randomness is dealt with to the extent presently possible by defining the three teacher models.

The third type of randomness entered the research process (intentionally, for proper scientific process) at the very beginning when all available vectors were randomly ordered (a necessary step, analogous to random sampling).

The scientific/philosophical intent is summarized by Occam's Razor. Attributed to William of Ockham⁴, Occam's Razor is the popular name for the principle of parsimony, which has a rich history predating and extending beyond the work and philosophy of William of Ockham, who popularized his predecessor Scotus' pronouncement: "*Pluralitas non est ponenda sine neccesitate*" or 'plurality should not be posited without necessity'" [16, 17].

This principle suggests that explanations for phenomena must not be made any more complex than would be sufficient. It is an ongoing debate as to whether complexity ought to be taken to mean that of a posited mechanism, or simply to mean the number of mechanisms posited. The former interpretation is that given two good explanations of the same phenomenon, the one that is less complex ought to be selected *when no other information is available*. The latter interpretation is the imperative to not make any more assumptions than is necessary in order to devise an explanation for a given phenomenon. These two imperatives are closely related, but there is room for disagreement. The debate may be settled when one considers that the less complex mechanism likely contains fewer parts, and hence fewer assumptions. That, in turn, is closely tied to basic results in Probability⁵, and shares a logical basis with Laplace's

⁴ Dating at least to William's predecessor John Duns Scotus, and possibly "back to Aristotle's statement in *De Caelo* that the number of postulates should be 'as few as possible, consistently with proving what has to be proved.'" [17].

⁵ The probability of a compound event is the product of the individual probabilities of the events making up the compound event. Hence, explanations for phenomena that have independent components (whose likelihoods are less than 1) have a resultant likelihood that is a product of the likelihoods of the independent components. This implies the chance of a lower likelihood for explanations with more components than for those with fewer. (Note that this is a Bayesian point of view, since *likelihoods* are assumed to be assignable to non-repeatable causes of phenomena, while *probabilities* require repeated experimentation.) The same goes for assumptions. Explanations with

principle of insufficient reason, which in turn can be thought of as the application of Occam's Razor to a choice among probability distributions. Laplace's principle, as interpreted by Kapur and Kesavan in their excellent explanation of entropy optimization, states that one should not select a more complex distribution than one has information to support [18]. The relevance of this principle is demonstrated in the conjunction fallacy, which reveals the universal lack of parsimony in human intuition [19, pp. 92–100]. Furthermore, Fost has linked parsimony to causality [20, p. 95], algorithmic complexity and pattern recognition, pointing out that the recognition of the broken version of a familiar figure (circle) as the complete figure is more parsimonious than interpreting the image as “two precisely aligned semicircles” [20, p. 59, 88]. This view that pattern recognition implements the principle of parsimony is supported by experiments in neuroscience [21]. (See Appendix D for more on John Duns Scotus & William of Ockham, Laplace, Newton, Einstein, parsimony, and the Lendaris–Stanley conjecture [22], as well as a brief discussion of arguments for and against the applicability of parsimony to technology.⁶)

The mere statement of a principle of parsimony is neither scientifically nor philosophically sound, and as such, Occam's Razor has been interpreted by many people to various conflicting ends [16, p. 266]. Rather than taking a broad sweep at Occam's Razor, or adding to the philosophical debates about it, this study offers a

dependent components are not considered here because they are likely to be circular arguments or exhibit inconclusive correlations (the four possibilities of correlation vs. causation), and are thus less favorable candidates for the ultimate explanation for any phenomenon.

⁶ Another expression of the connection among probability, Occam's razor and the Lendaris–Stanley conjecture is given in MacKay's explanation of “Why Bayes embodies Occam's razor” [148, p. 15].

scientifically sound way to test the applicability of the principle of parsimony to *technological* problems, or at least, to one branch of technology⁷. The probabilistic basis stated above leads to an interpretation founded on firmer mathematical grounds than some of the philosophical speculation that has been offered as to the basis and the use of this principle. The connection to Laplace’s principle of insufficient reason (which, to restate, advises scientists not to assume greater complexity for the causes underlying a *statistical* observation than there is reason to assume) serves to tie parsimony to the fundamental skepticism of the scientific method. (See Appendix D for skepticism as a precisely defined element of the scientific method, as well as challenges to the very existence of such.)

In this technological interpretation, an additional condition on Occam’s Razor is essential to its proper deployment: the choice of a particular explanation *given* that two explanations appear equally good based on available observations. Furthermore, as stated above, to validate its application to a technological dissertation, a technologically applicable hypothesis is needed.

The Lendaris-Stanley conjecture is such an application of parsimony to a field of technology, namely Artificial Neural Networks: Given two networks that can “successfully” learn a mapping, the network with the smaller *performance subset* (PS) will generalize better to previously unexperienced conditions [22].

⁷ For purposes of this study, technology is broadly defined as the human endeavor to develop devices, tools, or solutions to problems (*invention*), whereas science is taken to mean the endeavor to explain the workings of nature (living beings, environments, elementary particles, galaxies, brains, behaviors, and the like; i.e., *discovery*).

Prestructuring a multilayer perceptron necessarily reduces the performance subset (PS) of the network by disallowing, or removing, some of the network’s internal connections that govern the relationships between inputs (vectors, patterns, or questions) and outputs (classes⁸, categories, or answers). In other words, prestructuring reduces a network’s probability of settling at a solution that is different from the desired solution: All other factors being the same, the network’s chances of moving away from a good mapping are reduced by having a smaller set of possible mappings (a smaller PS) to search within.

For further motivation on prestructuring, consider that Computational Intelligence (CI) stands out among the engineering fields in terms of its share of heuristics in the design processes that are typically employed, as compared to the share of exact and rigorous theoretical underpinnings⁹. While engineering design is often “more art than science,” this characteristic is even more pronounced in the design of neural networks, evolutionary algorithms, fuzzy systems and their hybrid combinations than it is in similarly motivated subfields of Statistical Learning¹⁰. This dissertation aims to contribute to the balance of heuristics and rigor in the Neural Networks branch of CI by extending prestructuring to a new class of problems.

⁸ Both the term *class* and the term *category* are used in CI, but the former is more common. However, in this dissertation, *category* will be used exclusively (as “clave category, clave-direction category, samba category, etc.) so as to avoid confusion with the term “samba class” as in an educational setting. On the other hand, the term *classification* is standard in the CI literature, hence will not be replaced by the term *categorization*. Thus, for purposes of this dissertation, assigning *categories* to vectors is the act of *classification*.

⁹ This is not to say there are no rigorous theoretical underpinnings, but only that they share the justification for CI techniques with practical results and experiential heuristics.

¹⁰ Bös, for example, goes so far as to refer to “folkloristic tricks” [23] for the practical realization of neural-net designs.

The notion of *balance* permeates this dissertation. A significant component of the modeling performed herein is based on Information Theory, which is centrally concerned with the balance between entropy (randomness-related variance) and constraint (determinism, or lack of variance). Similarly, a balance is sought between the traditionally experience-based intuitive approach to neural-net design¹¹ and the rigorous requirements of statistical theory and its application to the extended (modern) scientific method (See Appendix D). This balance is sought in the determination of a structure in the problem space by information-theoretic modeling of the data at hand, statistical comparison of the models obtained, statistically sound practices of breaking the data up into training and test sets, and seeking statistical significance in the comparison between fully connected (traditional, standard) MLPs and prestructured MLPs¹².

One of several motivations for the use of domain knowledge in solutions to complex problems is the “*curse of dimensionality*” introduced by Bellman in his studies of

¹¹ How does one choose the interacting network parameters of momentum, threshold, random-number seed and learning rate, the time-varying changes imposed on each of those factors (parameter schedules), the strategies for pruning and growing, the choice of neuron type, the number of hidden layers or the starting size for each hidden layer, *when every such choice affects all the others*, and there is an infinite number of possible values to choose from? The designer must start somewhere, so heuristics based on experience or obtained from instructors, textbooks or software defaults are used, followed by the designer’s intuition as to how, when and how much to vary each parameter.

¹² According to the *Handbook of Neural Computation*, in the chapter on topology by Fiesler, “depending on the neural framework and learning rule, the term fully connected neural network is used for several different interconnection schemes [...] The most commonly used topology is the fully interlayer-connected one, where all possible interlayer connections are present but no intra- or supralayer ones. This is the default interconnectivity scheme for most nonrecurrent multilayer neural networks” [24, p. B2.5:1]. This is the sense in which “fully connected” is used in this dissertation. Fiesler continues: “A truly fully connected or plenary neural network has all possible inter-, supra-, and intralayer connections including self-connections. However, only a few neural networks have a plenary topology” [24, p. B2.5:1]. Topology, according to Fiesler’s definition, is a subset of the concept of paradigm, which includes element type and learning algorithm in addition to topology: “[T]he topology of a neural network consists of its frame or framework of neurons, together with its interconnection structure or connectivity” [24, p. B2.2:1].

adaptive control processes [25, *italics in the original*]. Bellman stated that “dense samples are hard to find in ‘high dimensions,’ In particular, there is an *exponential* growth in complexity as a result of an increase in dimensionality, which in turn leads to the deterioration of the space-filling properties for uniformly randomly distributed points in higher[-]dimension[al] spaces. [...] The only practical way to beat the curse of dimensionality is to incorporate *prior knowledge* about the function over and above the training data, which is known to be *correct*.” [26, pp. 211–2]. In the present research, the incorporation of prior knowledge “over and above the training set” is done through information-theoretic modeling of the training and validation data (which are all but the holdout set—see Section 2.1 on cross-validation) in order to extract structure from them¹³, and also in the ultimate selection among candidate models¹⁴.

Another argument for the indispensability of domain knowledge is that Statistics alone cannot determine the proper way to devise a research program: “Statistical techniques are most effective when combined with appropriate subject-matter knowledge. The methods are an important adjunct to, not a replacement for, the natural skill of the experimenter.” [27, p. 15]

Yet another motivating factor is given by Wolpert and Macready in their no-free-lunch theorem: “Given all possible search (optimization) problems, all generic

¹³ Other means of incorporating domain knowledge are also possible: One could poll experts as to how attack-point data should be parsed for clave-direction. Instead, the present work is based on information-theoretic means of identifying and codifying domain knowledge, making it a more generalizable study than the former option.

¹⁴ In addition to “running fit” on each candidate model (which gives detailed information about the probability distribution for a particular model, listing every input state and the corresponding conditional probabilities for every output state), each model structure is considered through the researcher’s musical intuition and to the criteria used in the classification process, and impossible or unlikely models are ruled out.

methods (such as genetic algorithms, neural networks, linear programming, and the like) perform equally in that for any class of algorithms that outperforms another in a set of problems, the latter is proven to outperform the former in all remaining problems *unless* aspects or features of the underlying function are identified and the algorithm is tailored with that information” [28]. This is a universal statement that prestructuring based on prior knowledge necessarily improves the performance of any search (optimization) over the generic version of that algorithm.

1.3 Objective of the Study

The research question for this dissertation is: For a real-world problem, can information-theoretically prestructured MLPs be shown to perform better in generalization (as defined by the Generalizing Ratio, Section 2.2.11) than fully connected MLPs that are otherwise the same?

“Otherwise the same” is intended to mean the same number of hidden layers, same number and type of inputs and outputs, same training algorithm and unit type.

The null hypothesis—the standard statistical alternative to any research thrust, which says that no effect will be discovered—is that both types of network will perform the same within statistical bounds. Rejecting the null hypothesis requires showing statistically significant generalization improvement at a target rate selected prior to experimentation.

Failure to reject the null hypothesis will suggest that parsimony is not a fruitful guiding principle for the application of Computational Intelligence to the cultural domain.

Rejecting the null hypothesis will provide evidence that parsimony is a valid guideline for the application of Computational Intelligence to the cultural domain.

1.4 Research Conducted

The quantitative target is to compare the Generalizing Ratio (GR, Section 2.2.11) values of prestructured and fully connected networks operating in the same environment—the same problem, the same input and output representation, and the same data. Neural-net generalization performance was expected to fall between that of an expert (ceiling) and a control (floor), but was actually shown to *exceed* the ceiling. One set of such real-world bounds is provided by benchmark experiments of supervised learning by human participants who were presented (a stratified, representative subset of) the same data as the “machine” portion of the research.

The floor value was provided by one group of participants with no prior experience of Brazilian-music performance: “naïve” human agents. This is one type of control used in the overall experimental design.

The ceiling bound was provided by a group of mid-level experts¹⁵ (non-Brazilian professional or semi-professional performers and teachers of Brazilian music).

A further randomized control is supplied for the prestructured networks by a randomly (in fact, haphazardly) prestructured network known as the “lousy control structure.” Since prestructuring is a way of removing resources, an uninformedly,

¹⁵ Mid-level experts were consulted for this stage for scientific rigor: The high-level experts available were already consulted to guide the I/O mapping for the study. To avoid the “blunder committed many times in published papers in top[-]rank journals” [29, p. 247] of basing assessment work on the same data as development work, different experts had to be consulted for the mapping and the benchmark.

haphazardly prestructured network must perform less well than any properly designed network. This is an exaggerated but related setup to having a group unknowingly receive placebo in medical research.

The main comparison is between the generalization performances of a fully connected (standard) multilayer perceptron (MLP) design and that of an information-theoretically prestructured (biased) multilayer perceptron design.

Standard model evaluation with k -fold cross-validation is used to determine the size of the fully connected MLP. Given this hidden-layer size (also called “network size” [29]), a number n of MLPs are trained and tested. Each of these n networks is initiated using a different random-number seed, and is trained from scratch. Testing is done on the holdout data set that was removed from the design data and put aside at the conclusion of data preparation.

The random-number seeds are listed in Appendix O.

The design of prestructured MLPs is achieved through information-theoretic modeling of the problem space (minus the holdout data) using Reconstructability Analysis (RA) as implemented in *Occam3* [30]. Testing is, again, done on the holdout set.

While the above comparison is the primary goal of the proposed and approved dissertation research, an additional type of generalization performance is included in the comparison for added rigor and practicality of the results. This last type of generalization data comes from RA generalization, as Reconstructability Analysis is not only an analysis method, but also a form of Machine Learning.

The proper statistical technique for making the desired multi-way comparisons is the ANOVA (analysis of variance) with the Dunnett, Bonferroni, and Tukey methods. The justification for this is given in Chapter IV (Methodology) and Appendix D on Statistics.

As an additional source of information and insight, various two-way comparisons can be made between the fully connected and prestructured network types' generalization performances.

The overall research, then, consists of problem definition, data acquisition, problem-domain investigation, neural-net design, information-theoretic modeling (RA implemented in *Occam3*), determination of statistical power and significance targets and the repetition of experiments to achieve these targets, and the final comparison(s).

1.5 Contributions

- Ranking the utility of RA model-search and model-evaluation criteria (with BIC as the best for the current application, followed by Information)
- Evaluating RA-search heuristics (all-model bore-expand versus mixed-model bore-rotate, with results in favor of the former)
- Establishing a methodology for applying the results of Reconstructability Analysis to neural-net prestructuring (biasing) by reflecting model structure in network structure (see Section 5.1),

- Demonstrating enhanced generalization ability in three types of prestructured neural networks over fully connected neural networks, RA itself, and human performance,
- Establishing the technical and theoretical foundation for providing the music industry and the Afro-Brazilian music community with a useful tool to aid in musicians’ and students’ internalization of the clave construct as interpreted in the traditional Brazilian idiom.

1.6 Significance and Rationale

Occam’s Razor, also known as the principle of parsimony and closely related to the principle of insufficient reason, is one of the fundamental tenets of the “scientific method.” (For aspects of the debate on whether such a thing exists or not, and how this is relevant to the present work, see Appendix D.) It has the following consequence for search (optimization) and modeling problems: The smallest, simplest model that performs satisfactorily should be preferred over any model that is more complex (whether in terms of the number of parameters, i.e., model size, or the nature of “fitting,” i.e., linear versus nonlinear).

Two difficulties in applying Occam’s Razor successfully in search and modeling are that (1) the term *satisfactory* cannot be rigorously defined (as is the case with “successfully” learning a training set [22]), and (2) performance cannot be known in advance.

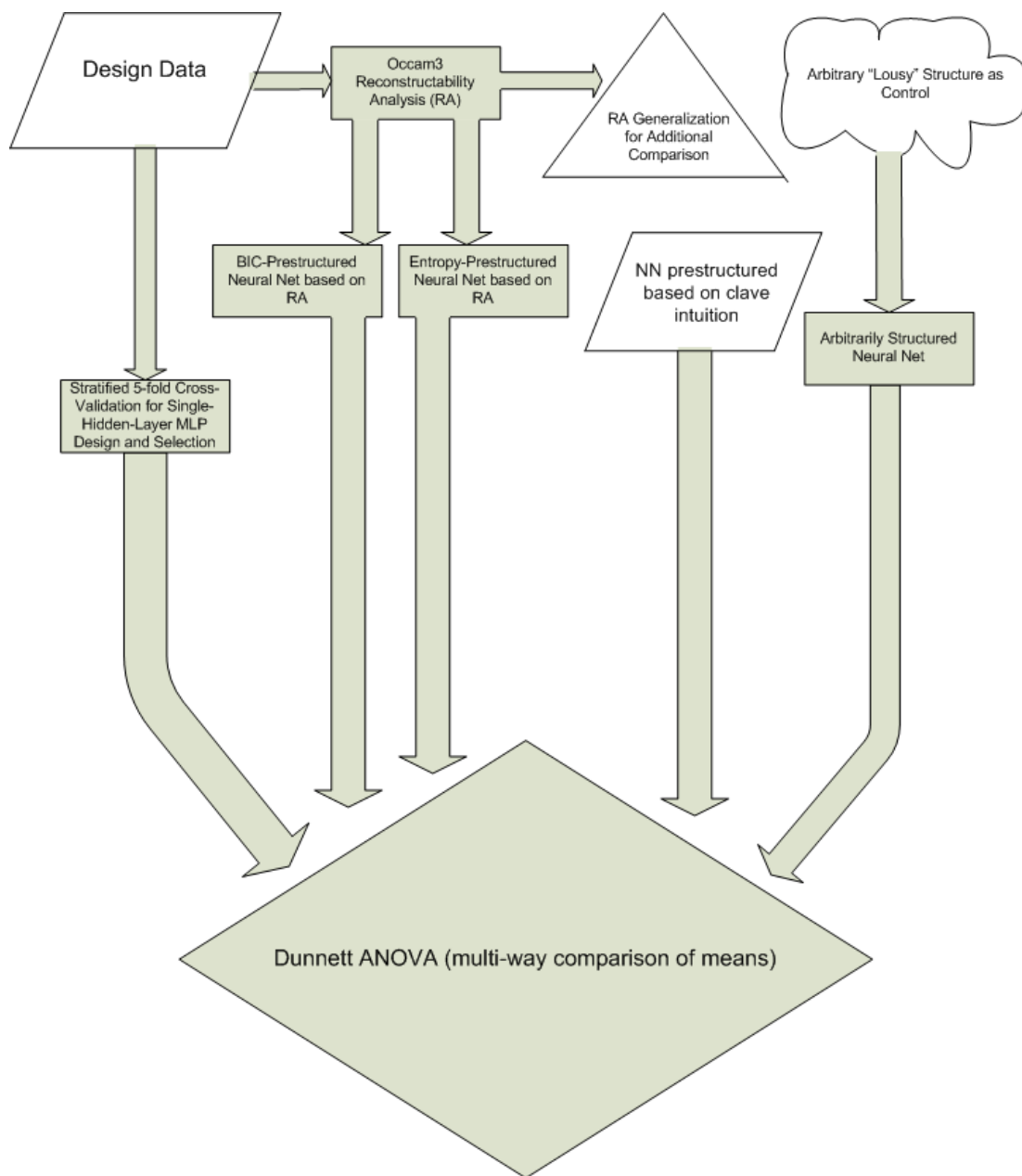


Figure 1: High-level block diagram of overall research flow, showing the design data, the structures compared for their generalization performance, and the primary comparison technique (ANOVA). Note: There are more than two RA-based prestructured neural networks, as explained in the section on RA search heuristics (Section 1.16), but only two were “finalists” based on their performance in the preliminary experiments. Also, the shading is an artifact of Visio, and not meaningful.

Is a solution that performs 90%-correct on an out-of-sample test set (holdout set) satisfactory? How about 99%? What if there is another solution that would have performed at the 99.98% level?

Furthermore, since the out-of-sample test set is not available during the design phase, no choice of model or algorithm can be fully justified as to its eventual performance. (This is in the nature of performing scientifically rigorous research.)

Instead, one must rely on statistical methods of *estimating* future performance. Many such studies exist [31–36, etc.] and support the use of prior knowledge (domain knowledge) for biasing or tailoring optimization algorithms, a particular form of which is the information-theoretic prestructuring presented in this dissertation. The existence of support for this idea does not imply the existence of methodology. Hence, developing the path from domain knowledge to its use in prestructuring is the primary contribution of this study.

1.7 Assumptions and Hypotheses

The two fundamental assumptions underlying the present research are that multilayer perceptrons (MLPs) whose connections are limited to a subset that reflects the underlying structure of the problem to be solved will generalize better than fully connected (in a sense, brute-force) ones, and that a Computational Intelligence agent can be devised to capture the culturally specific meaning encoded in a timing/accent pattern that carries a sense of clave direction.

A requirement for generating data apropos of this research is that a human expert can conclusively perform the classification of timing/accent patterns whether

they are presented in written (coded) form or in audio. Such data are used for training and evaluating the performance of MLP-type neural nets in the various stages of the research.

1.8 Scope, Limitations, and Constraints

The scope of the research is the comparison of generalization performance in fully connected and prestructured MLPs. Other types of neural networks and other forms of Computational Intelligence (except for RA classification) are not investigated.

The practical reason for this is that the design of even a single multilayer perceptron, when undertaken with rigorous completeness, is an infinite task. There are no precise theoretical bounds guiding the designer in his/her choice of a variety of parameters and design decisions, from the number of hidden layers to the type of activation function, to the number of elements per hidden layer, to the output encoding (one-up, binary, or others), to the learning schedule and the further variations the learning schedule can take on for each of its target parameters (momentum, derivative offset, step size, etc.), to the selection of training data, to pruning options and early stopping, to the techniques of performance estimation.

Hence, the design of one MLP is potentially an infinite task, and while infinity is not practical, it is still necessary to make every reasonable attempt to achieve generality and rigor. Therefore, the focus of the present research is only on multilayer perceptrons and information-theoretic prestructuring for generalization improvement, with a limited list of practical neural-net design practices allowed for all models under comparison.

The musical aspect bookends this research in that it is the source of data for experimentation (the initial role) and a guide for practical decisions in design so that the results may be used to develop a useful musical-training product (the final role). In addition, musical intuition is used during the research in the vetting of information-theoretically derived models.

Additional necessary limitations come from the types and use of data. In order to avoid an avalanche of factorial experimental combinations, one primary path was selected for each of the following experimental forks-in-the-road:

- The teacher-model context
- The type of holdout set
- The clave-direction output sought and the encoding of said output
- The number of hidden layers for the fully connected networks
- The number of processing elements per hidden layer
- The search and evaluation criteria for Reconstructability Analysis, and
- The choice of output threshold for interpreting the neural nets' classification results.

The explanation of the choices made follows.

1.9 Teacher Models

The data available for this research comes in the form of three “teacher models” or contexts: the strict teacher, the firm teacher, and the lenient teacher. These models are based on real-life interactions with accomplished, world-famous teacher-performers of Rio-style samba percussion and song. In order to have a classification problem that is both well-defined and not trivial, the firm-teacher model was selected. The musical interpretation of the firm teacher is one who is not fanatical about adherence to the *partido-alto* form *per se*, but is nonetheless wholly concerned about the

preservation of the *carioca* interpretation of its clave direction in performance and teaches to that standard. Thus, only data that is classified according to the firm-teacher model is used in the present research. (Data were also classified according to the other two teacher models. This was necessary for developing and refining all three models. It also provides additional data sets for future work and other studies.)

1.10 Selection of Holdout Data

Two types of holdout sets were created at the early phases of the research. Although the holdout sets overlap—and thus the design sets also overlap—the isolated computerized nature of Computational Intelligence research makes it possible to keep the two research paths from contaminating one another’s data. This creates the possibility of two different takes on the crucial notion of holdout for performance estimation (cf. Section 2.2). One holdout set is the weak holdout. (The meaning of a “weak vector” is explained in Section 4.8.) This holdout set contains only weak and very weak vectors, so the training and validation sets (which together comprise the design set) contain all the average, strong and very strong examples. The reason for the weak holdout set is that in real-life music training, individuals are taught using the best examples of the tradition in question, not the weakest (most esoteric) examples. The weak holdout set mimics this situation by using the more typical examples of timing/accents patterns for training and validation (design).

The other holdout option is the standard scientific (randomly selected) holdout set. The type of random holdout was selected as the primary path because while musicians are usually trained with typical (“strong”) examples, with years of practice,

they do generalize to somewhat less obvious cases, but usually with decreasing *consistency* for more unusual and rare patterns. (For a student or even a performer of traditional music, this is acceptable because very atypical patterns are played very rarely.) The present author took several years to derive a complete and consistent theory that explains (from a *partido-alto*-based *samba carioca* point of view) the clave direction of even the oddest (in a sense, “noisiest”), most esoteric patterns. Although one cannot expect all other musicians to be absorbed in the same type of focused analytical study, one does indeed expect such high levels of performance from a computational system. Since any resulting music product must be able to handle the most difficult cases, the random holdout was deemed preferable for the primary experiments to the weak holdout.

1.11 Output Encoding

The following output encodings were considered and experimented with in the investigation stage: The true binary output encoding (two bits to represent the four clave-direction categories), the true one-up encoding (one dedicated “wire” per output category), a three-output encoding based on the observation that one of the four categories is complete opposition to clave sense (and hence can be represented by all outputs going low), the “one-output-at-a-time” (OOT) encoding (which considers only one of the four clave-direction classes at a time), and other encodings that take membership degree into account in the definition output categories. The encoding selected for primary experimentation is “one output at a time.” This means that the networks only ask whether a pattern belongs to a particular clave direction or not. If it

does not, the way in which it falls outside that category is left until a later stage in the modular design of the intended musical product.

1.12 Number of Hidden Layers

The number of hidden layers is selected such that the comparison with prestructured networks would be of “apples to apples.” The prestructured networks have a single hidden layer, so the fully connected networks to compete with them must have the same type of resource allocation in the form of hidden layers. (More than one hidden layer in multilayer perceptrons is believed to help discriminate among multiple classification criteria to be learned.)

1.13 Number of Hidden-Layer Elements

The number of elements per hidden layer (sometimes called “network size”) is determined through an experimental procedure called k -fold cross-validation [29]. In this method, the generalization performance of any particular instantiation (a network with a certain number of hidden-layer elements) is compared with the performances of various other instantiations. (The only difference distinguishing all the network instantiations is the number of elements in one hidden layer; all other parameters are kept constant for this experiment.)

The question arises as to which two network sizes ought to be compared first, and how much should network size be altered for subsequent comparisons. Hence we see that not only is the determination of network (design) parameters heuristic and

cyclical in NN design, even methods of systematically determining parameter values require heuristics, guidelines and experience.

One possible route is to choose a very small and a very large network size, and perform a binary search¹⁶. However, binary searches are optimal (i.e., expected to succeed more than 50% of the time when performed an infinite number of times) in the absence of experience about the value sought. In the case of the Neural Networks field, the recommended search starts with a comparison of 5-hidden and 25-hidden networks [37]. If the performance is close (and good), or if the next larger/smaller network does not improve on these two, then binary search is employed. If the performance of the 5-hidden and 25-hidden networks is significantly different (which is not precisely defined), then the network size is varied in the direction of the better-performing size. For example, if the 25-hidden network performs significantly better than the 5-hidden network, a 50-hidden network is built. “At this point we compare the performance of the networks with 25 PEs and 50 PEs in the hidden layer. If the network with 50 PEs in the hidden layer outperforms the network with 25, we would build another network with, say, 100 PEs in the hidden layer. We continue this building and testing cycle until we build a network where the performance on the network with fewer PEs outperforms the network with more PEs.” [37]

¹⁶ Binary search is the most efficient method for guessing a number between two limits. The first guess is the middle of the range. If too high, the next guess is halfway from the last guess to the low end of the range. Thus, the range is progressively shortened, and each guess brings the range of possibilities down to half the previous range. This method is optimal in the case of complete randomness or ignorance. When the value sought is not random, the process can be aided by experience. Most classification problems have optimal hidden-layer sizes either between 5 and 25, but with the possibility of going down to 1 and up to 50, or between the numbers of input and output elements. Neither of these heuristics is a guarantee, however.

This method was used to determine the best network size for the fully connected network (using only the default NeuralWare NeuralWorks initiation—tests of generalization performance between fully connected and prestructured networks require multiple initiation seeds determined randomly). Ideally, one would also use multiple seeds for the design stage, but following such ideal practices leads to the ∞^∞ number of combinations that can be sought in neural-net design.

The network size for the fully connected nets was determined to be 3-hidden.

1.14 Criteria for Model Evaluation and Selection in RA

The search and evaluation criteria for Reconstructability Analysis as implemented in *Occam3* are Information (see Section 2.5), α (chi-squared, or χ^2 , significance for the full change from reference to model), incremental α (from last model to current model), Akaike’s “an information criterion” (AIC), and Schwartz’s Bayesian information criterion (BIC) and percent correct on test. All of these criteria have trade-offs with respect to one another¹⁷, and model searches with each one under otherwise similar conditions lead to highly different models. (Part, but not all, of the reason for this is the *curse of dimensionality*).

¹⁷ In RA, complexity is defined in terms of degrees of freedom, without regard to functional form. According to this, the equations $y_1 = a_1x_1 + a_2x_2$ and $y_2 = \sqrt[a_1]{\sinh(a_2x_1)^{x_2}}$ are equally complex (two degrees of freedom). This is why RA is used for nominal data rather than regression. A variant called Fourier RA is used for regression. In using AIC and BIC within RA, a further trade-off arises. AIC has a penalty term that is twice the number of parameters in a given model. BIC includes the sample size in its penalty term with less emphasis on the number of parameters. Thus, when certain statistical assumptions (beyond the scope of this dissertation) are not valid, BIC can *underfit* [37, p.2].

Not considered for model evaluation here is percent correct on test. This is RA's equivalent to generalization-performance estimation, and is only used for the final comparison once all "hypotheses" (networks and RA models) have been tested.

The remaining criteria fall into two groups: those that can be used for model selection and those that are used for accepting or rejecting a model once it is selected. The latter are the two α values. No model selected by any other criterion is admissible unless its α value suggests that the likelihood of being wrong in claiming this model is different from the reference is very small (less than 5%). Hence, the α criteria are not used for model selection during the RA search, but used afterwards by the researcher to check the statistical admissibility of selected RA models.

The remaining criteria are AIC, BIC and Information. AIC and BIC are the two classic penalty functions for model evaluation. Since their development, many alternatives have been suggested for various specialties. These include AICc (AIC-corrected), CIC (curvature information criterion, based on AIC), NIC (network information criterion, a variation of AIC for direct application to neural networks¹⁸), TIC (Takeuchi's variation of AIC) and others. Some discussion of these is given in the relevant section on statistical background.

AIC is an estimate of the expected value of the Kullback-Leibler distance between a given model and the *unknown* underlying distribution [38, pp. 1–2]. BIC is a similar measure, and they both include a term that estimates the model's fit to the inferred underlying distribution. Furthermore, they both include a penalty term that

¹⁸ As an alternative to cross-validation for deciding the number of hidden-layer elements, not for prestructuring the network.

reduces a model’s final score based on its complexity. Hence, AIC and BIC (and the other “IC”s) are implementations of Occam’s Razor for tempering degree of fit.

The main outward difference between AIC and BIC is in the choice of penalty terms. In AIC, the penalty is a function of the number of parameters in the model [38, p. 2] (similar to degrees of freedom). In BIC, the penalty is a function of sample size *and* degrees of freedom, with the main emphasis on sample size [39, p. 233, Eq. 7-35, and 28, p. 215]. This is a common trade-off between statistical and information-theoretic metrics. For example, Transmission (Section 2.4) is the magnitude of difference between models without concern for statistical significance whereas the χ^2 likelihood ratio is the statistical significance of the difference without regard to its magnitude.

“For model[-]selection purposes, there is no clear choice between AIC and BIC” [39, p. 235]. (In fact, Rodríguez claims that under certain geometric conditions, both AIC and BIC choose the wrong model 100% of the time! [40, pp. 4–5] On the other hand, Stone has shown that AIC is equivalent to leave-one-out cross-validation (LOO) [41], which in turn has been shown by Zhang to overfit [42], which is consistent with our expectations of AIC¹⁹.) Kuha adds that AIC and BIC’s “success in consistently selecting good models for observed data ... is a much more complicated question both to ask and to answer” [32, p. 216–7]:

Some of its difficulty lies in defining what is meant by a ‘good model.’
[...] AIC and BIC represent two rather different answers to this

¹⁹ The Statistics literature further shows that k -fold cross-validation is superior to any other form (thus including LOO) [30], and that stratified cross-validation is superior to holdout, LOO, Bootstrap or any non-stratified k -fold cross-validation [41, 42]. Cross-validation, however, is a technique that can be used *with* Reconstructability Analysis, but not as part of the search methodology of *Occam3*. Hence, the task of choosing among Information, AIC and BIC remains.

question. The aim of the Bayesian approach motivating BIC is to identify the models with the highest probabilities of being the true model for the data, assuming that one of the models under consideration is true. The derivation of AIC, on the other hand, explicitly denies the existence of an identifiably true model and instead uses expected prediction of future data as the key criterion of the adequacy of a model. [32, p. 216–7]

From this it would seem that AIC is the correct choice of RA criterion since its motivation for derivation—its very reason for existing—is the same as the RA portion of the present research: to find the best predictor of future data.

However, Kuha continues to list problems with the performance of AIC, such as its well-documented tendency to select models too large for generalization success when data are scarce [32, p. 217] as well as its decisive success over BIC “in large-sample simulations in which the true model was not included” [32, p. 218].

The task remains to identify the right set of conditions relevant to the present research, and then to choose among AIC, BIC and Information (which is kept in the mix because of the potential for failure in AIC and BIC, as well as its purely information-theoretic, non-statistical nature).

The decision between the two information criteria depends on two factors: dataset size and number of model parameters. Since the latter varies in the course of each RA search, and since the former varies during the course of the present research, even known asymptotic behaviors can only be vague guides. The ideal course of action is to carry out every RA search once with AIC as the criterion, once with BIC, and as will be discussed, once with Information.

Having done nearly all of the early investigative RA searches this way, in the interest of being able to derive a definitive methodology for neural-net prestructuring, the author has endeavored further to find an answer to the admittedly open question of which criterion to search on. Before reporting on these further explorations, summaries of the theoretical pros and cons of AIC and BIC follow (in the next two paragraphs).

The two main problems with AIC are that AIC underpenalizes as sample size approaches infinity [39, p. 235; 32, p. 222], and that it also underpenalizes when the sample size is so small as to be about three times the number of parameters [45]. The latter is not a concern in the present research because it implies a sample size on the order of 50. None of the data sets used here are that small.

While BIC is proven to be asymptotically consistent (chooses the best model when sample size equals infinity), it is known to underfit if the underlying reality generating the data is not of a finite-parameter nature [38, p. 2]. Since the present research is based on the premise that greater complexity reduces the probability of achieving a target mapping, and that BIC penalizes more heavily for model complexity, BIC appears to be the best choice for the model-search criterion, but it also frequently overpenalizes in favor of too simple a model [32, p. 224].

To make the final decision, these theoretical foundations are compared with an empirical study of correlations between RA search criteria and the onset of overfitting. Three types of RA runs were performed on *Occam3*: loopless searches, depthwise (width-one) all-model searches, and wider all-model searches informed by the depthwise searches. Plots of percent correct for training and test data were created with respect to

model degrees of freedom²⁰, along with one or more information-theoretic metrics that may aid in the identification of the onset of overfitting.

Figure 2 shows overfitting in action with percent correct on test within RA. The percent correct for training and test data track each other up to some level of model complexity, at which point test accuracy flattens out as training-data accuracy continues to increase. The latter is the “optimism of the training error rate” [39, p.228]. Eventually, test-data accuracy not only flattens out, but also exhibits a downturn. The model clearly overfits the data from this point on, and will thus be unable to generalize well to examples not encountered during training.

For the same loopless search that generated this example of overfitting, information-theoretic criteria calculated by *Occam3* for these models are plotted along with the training- and test-data accuracies in the following plots. It is clear that some metrics are better indicators of the onset of overfitting than others.

²⁰ Note that this is not the typical x axis for an overfitting plot. Degrees of freedom (model complexity) is used in place of training time. Model complexity in an RA search is, in fact, an analogue of the number of training passes in neural-net (or other CI) training. An RA search does not alter a given hypothesis (a neural net, a genetic algorithm, etc.) over time, but moves from model to model, monotonically in the direction of greater complexity over time.

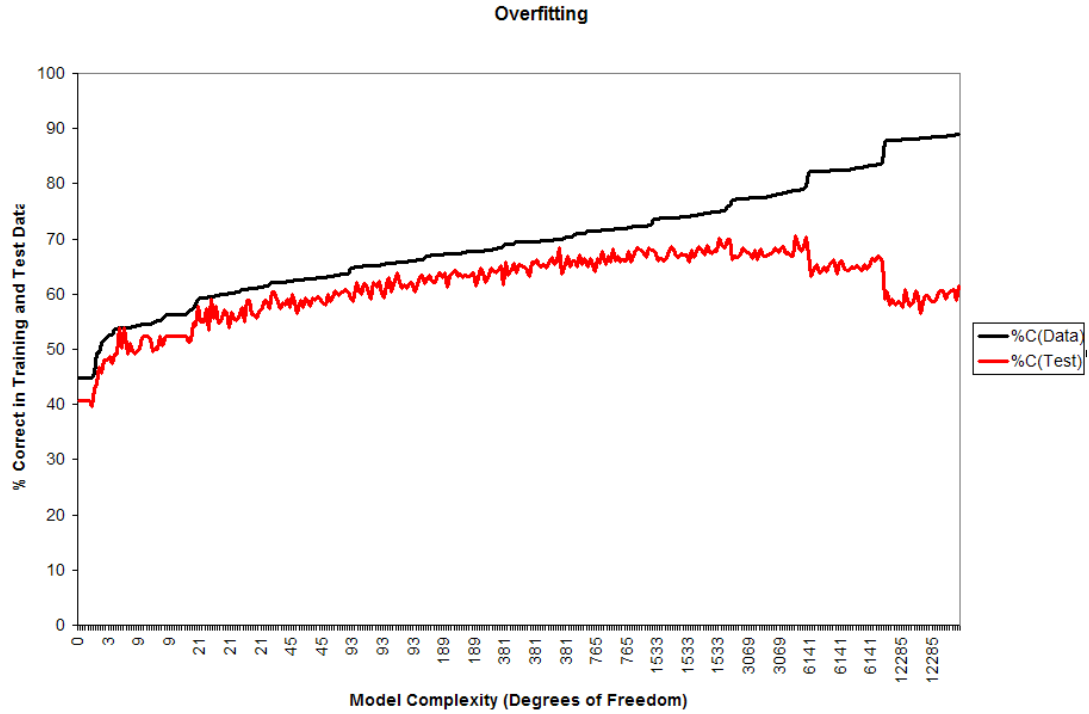


Figure 2: Overfitting in terms of model complexity, an experimental reproduction of the classic plot.

An order of magnitude leap in the difference in degrees of freedom between each model and the reference model is observed at the point where the test-data accuracy goes from flattening out to noticeably decreasing (Figure 3). This is plotted against model level, which is the number of variables in the model. For this study, model level and degrees of freedom are closely related since degrees of freedom have to do with the cardinality of variables in the model, corrected for overlap, and not with the nature of the relationships among the variables [46, 47].

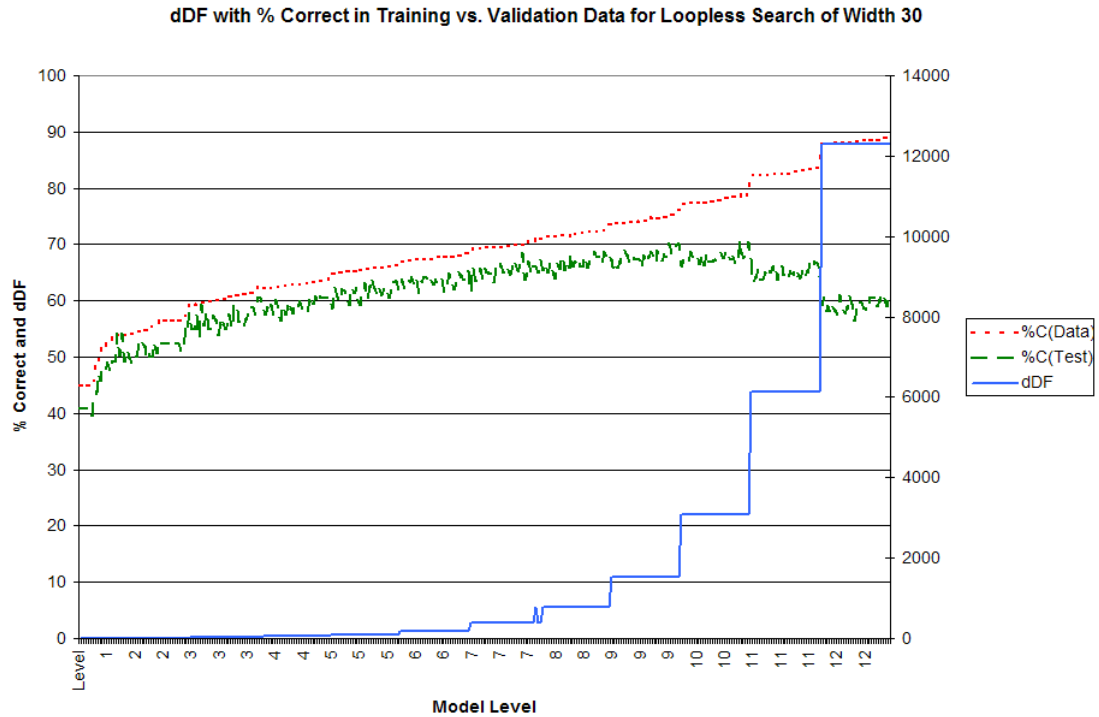


Figure 3: Change in degrees of freedom versus model "level." Level, also related to model complexity and search time, is the number of input variables in any single-component model, where a component is a collection of variables related in terms of their joint effect on the output.

When training and test are tracking, the numbers of degrees of freedom are between 0 and about 500. As the two begin to split, degrees of freedom are in the range of 1000 to 3000. When the percent correct on test shows a downturn, the degrees of freedom leap first to 6000, then to over 12000.

It would appear that large changes in the degrees of freedom are good indicators of the onset (and growth) of overfitting. This finding supports the application of Occam's Razor to technological problems: large gains in model complexity lead to greatly diminished generalization ability (as overfitting is known to bring about).

Another promising indicator of overfitting is the significance level α (Figure 4), which states the upper limit for the probability of a type-I error, where the null hypothesis rejected by such an error is the reference model. When α is equal to one, any model at that, or higher, complexity has a 100% chance of attaining the stated training-set accuracy by pure chance alone.

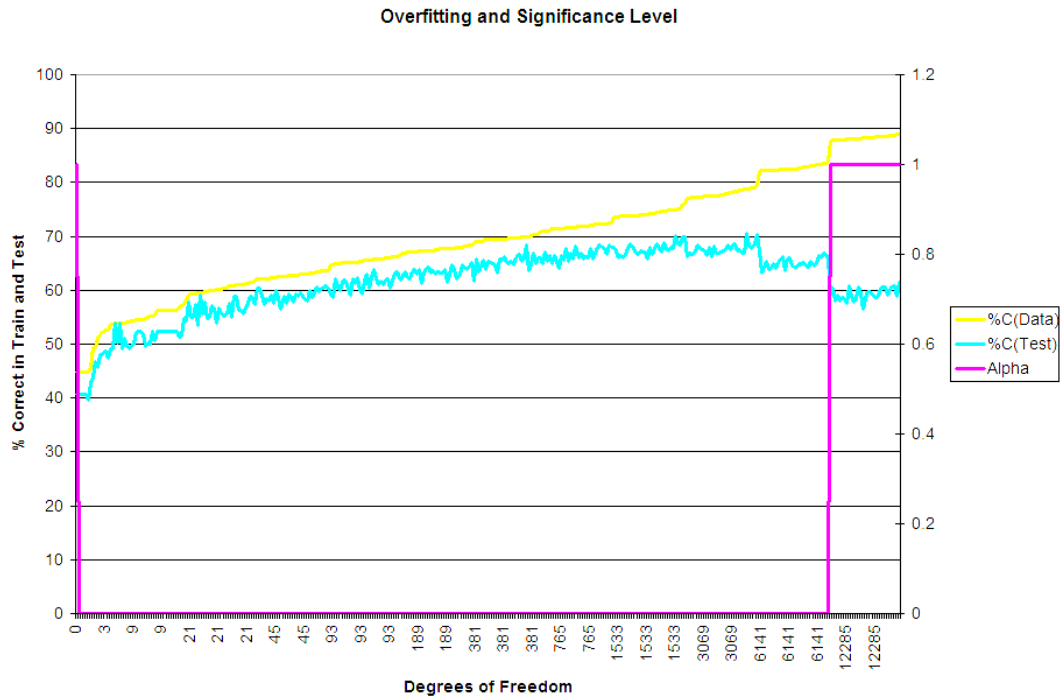


Figure 4: Overfitting and cumulative significance level

Less helpful is model Entropy (Figure 5). The change in Entropy (also called Uncertainty) does not appear to give a very clear indication of overfitting (Figure 5), as it increases near-monotonically for increasing complexity, but a close inspection reveals that, in this case, the magnitude of jumps in the value of Entropy increases in concord

with the mismatch between training and test accuracy. If this type of apparent correlation is observed with other types of RA searches, Entropy may be a lesser indicator of overfitting, requiring an examination of its trends rather than providing a threshold value.

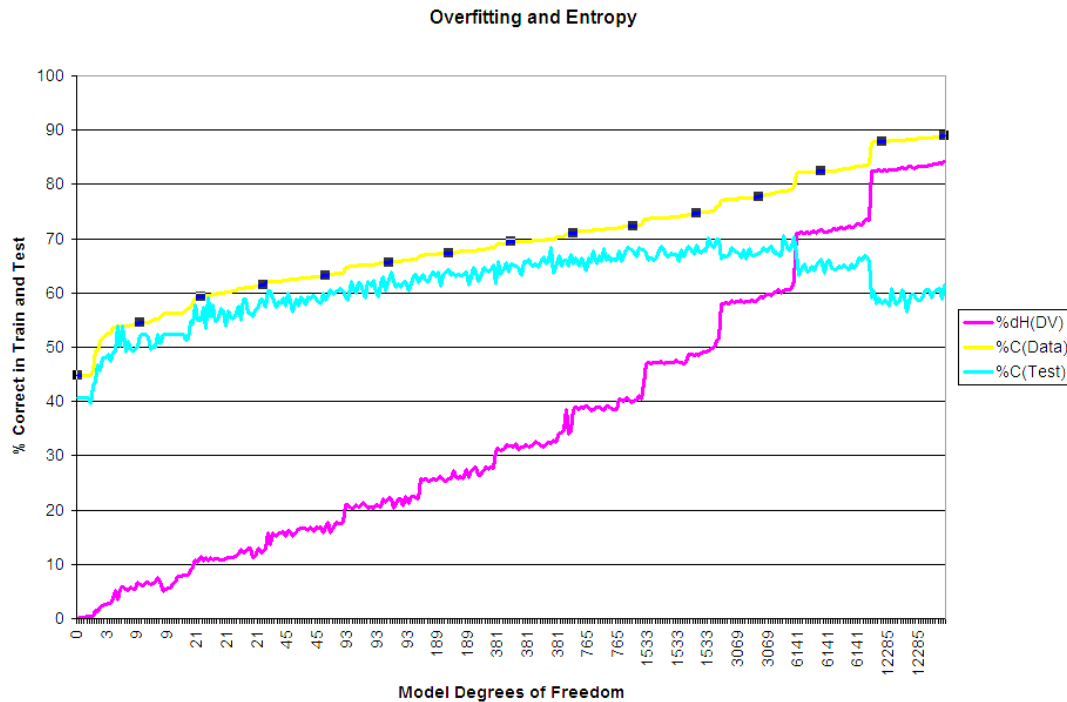


Figure 5: Overfitting and change in model Entropy

The final plot for loopless searches (Figure 6) concerns AIC and BIC. Note that what is plotted is the change in AIC (dAIC) and change in BIC (dBIC) from model to model. The best model (according to either criterion) has the minimum value of AIC or BIC, which means the maximum dAIC or dBIC. In this case, the cutoff points recommended by the two criteria are at the level of complexity at which each one reaches its maximum value. (Note that dAIC and dBIC values are on the right vertical

axis.) Figure 6 shows AIC reaching its maximum value at or near the point where test-data accuracy begins to flatten out. In comparison, BIC is too conservative to be used as an indicator of overfitting, as it reaches its maximum an order of magnitude too early in degrees of freedom. AIC appears to drop just in time to signal the diminishing returns of overfitting.

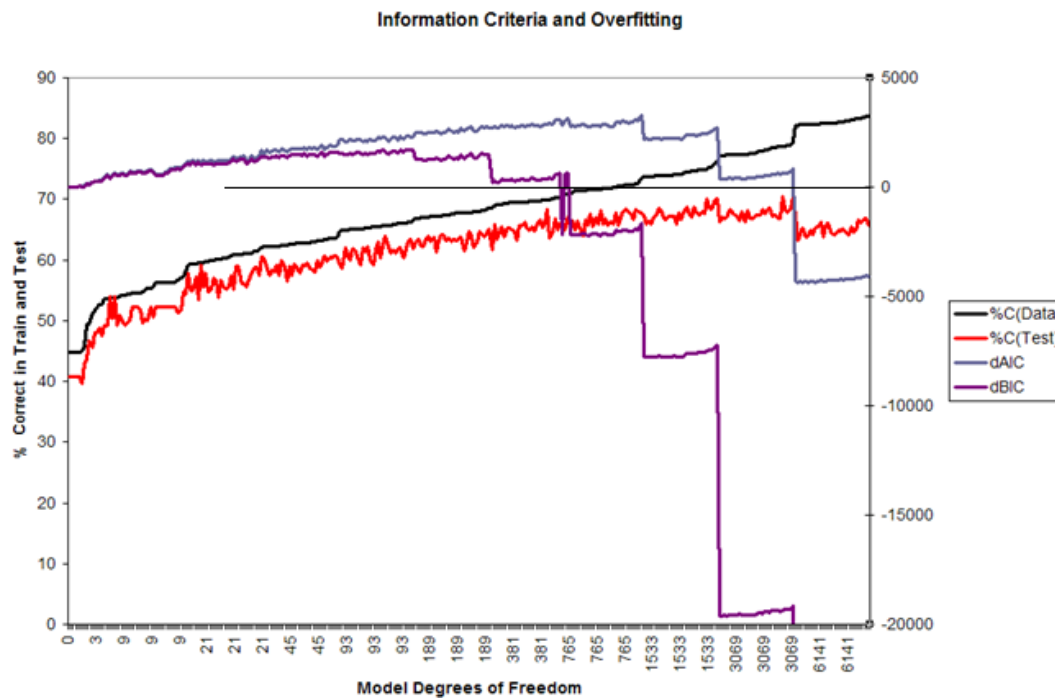


Figure 6: Change in AIC, change in BIC, and overfitting: A straight line has been added where $dAIC = dBIC = 0$, showing that BIC may indicate the beginning of the training/test performance split, but that AIC clearly marks the drop in test performance.

This plot is consistent with the theoretical understanding that BIC penalizes more heavily (or conservatively) than AIC. However, the exact nature of the difference is important: BIC here appears to over-penalize in that it may be quitting too soon. On

the other hand, one could say that AIC waits too long (until degraded test performance is visually evident), and might not be implementing Occam’s Razor sufficiently²¹.

Furthermore, this is not the final word on criterion selection because all the prior plots were for loopless searches, and the bulk of the modeling is done after the loopless searches during the all-model searches.

In the next plot (Figure 7), we see that for all-model searches AIC, BIC and Entropy are useless in indicating the onset of overfitting although each begins to flatten out. None, then, is conservative enough. The plot in Figure 8 indicates that a similar lack of decisive indication is seen in Entropy change and Information as well.

While not at all clear indicators, trends in Information, AIC, BIC, and the change in Entropy may still be able to *signal* the oncoming occurrence of overfitting through their first derivatives. The change in Entropy for more and more complex models increases at a lower rate (slope) for models with overfitting than it does when training and test sets track each other with respect to model complexity.

The most promising candidates for information-theoretic prediction of the onset of overfitting appear to be dDF , α , $dAIC$, and Information. (This is not much of a gain over where this inquiry started, but this is the finding, and thus must be considered.)

²¹ Within the realm of RA training and recall, Shervais and Zwick observe that “models picked by ΔBIC do better on generalization (test or recall data) than the more complex models picked by ΔAIC ” [46, p. 533]. Hence, the score continues to be even, and appears more and more conditioned on the application and approach.

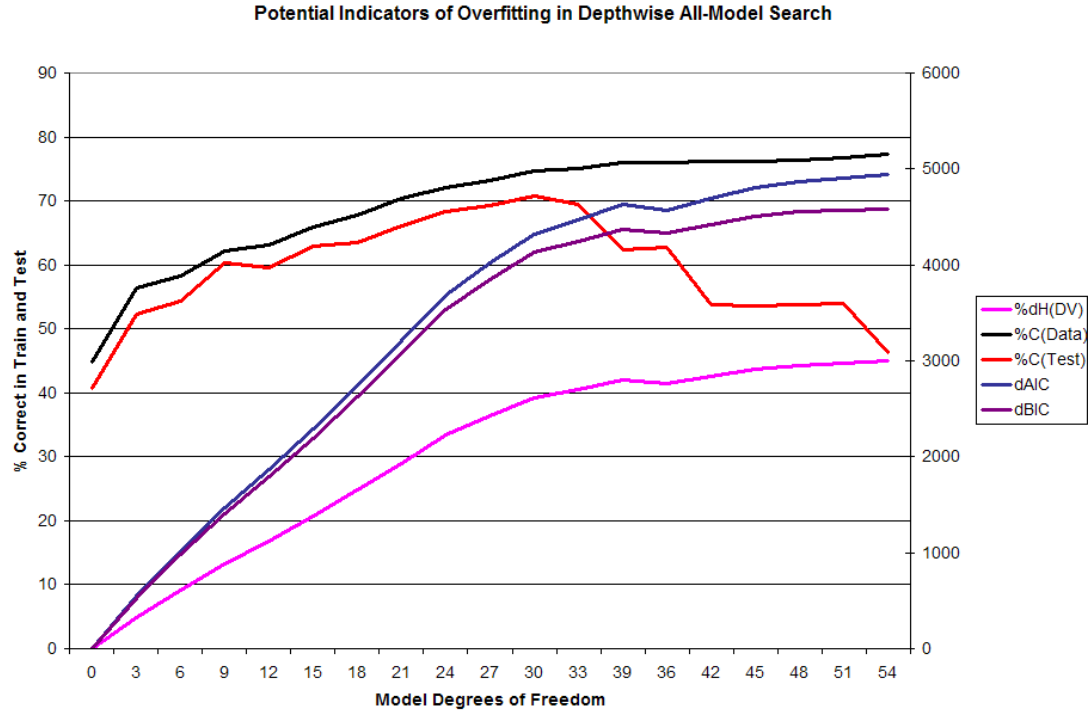


Figure 7: AIC, BIC, percent change in Entropy (Uncertainty), and overfitting in an all-model search

Among these, dDF (change in the number of degrees of freedom) may be used in a confirmatory mode, as has already been suggested for α . In other words, each model selected by whichever other criterion must pass the test of $\alpha < 0.05$ and dDF less than 50% of the range in that particular search.

Although AIC was a better indicator for loopless searches, since both criteria failed to catch the onset of overfitting in all-model searches, it would only make sense to at least use the more conservative option (BIC) for those situations. Hence, it may be best to use AIC for loopless searches and BIC for all-model searches (if one were to use these criteria at all).

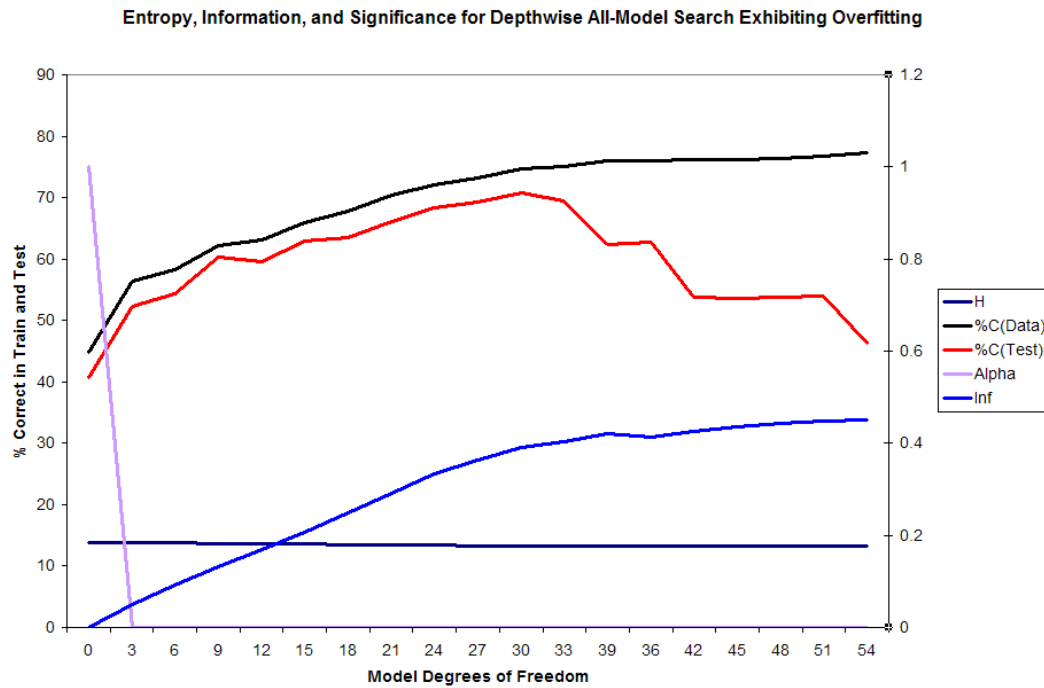


Figure 8: Entropy, Information, significance level, and overfitting in an all-model search

The final decision between the statistical penalty metrics and Information is one that neither the researcher's experiments, nor his class notes, nor the literature seems to give any *conclusive* answer to. Therefore, the choice must be to use both types of criteria²²: Information for information-theoretic model evaluation, and BIC and AIC (as indicated immediately above) for statistical model evaluation.

²² It was found, after writing this statement, that this is precisely the conclusion reached by Kuha: "an approach of using the two criteria together (as well as significance tests and perhaps other indices of model fit) has been advocated here. When the criteria agree on the best model, this provides reassurance on the robustness of the choice. Even disagreement usually rules out many models and

This necessity will result in two sets of prestructured neural networks, one according to RA searches by Information (tempered by α and dDF) and one according to RA searches by AIC-followed-by-BIC (also tempered by α and dDF).

For the reasons given above, and as explained below in Section 1.16, AIC is used for the first two steps of the *loopless bore-rotate search*. BIC is used for its last step and for the *all-model bore-expand search*.

1.15 Output Thresholding in MLPs

The choice of output threshold arises both from the operation of the NeuralWare NeuralWorks software package and from the nature of MLP outputs. The output nodes of a multilayer perceptron typically deliver numerical values in a predefined range. Whether using a binary, one-up, OOT, or other output encoding, a final decision must be made as to whether to accept such numerical values as ‘yes’ or ‘no’. In the present research, this is achieved by selecting a threshold above which any value is a ‘yes’ and below which any value is a ‘no’. The selection of this threshold, however, must not be arbitrary; it must be intelligent and meaningful to the problem at hand.

Two subsets of data were created (as test sets) to help with the process of intelligently choosing a threshold for interpreting neural-network outputs: a file of well-known traditional samba patterns that any end-product must correctly identify, and a file of 154 vectors from categories 1 and 2 (FWD and REV) noted for their

provides bounds for the set of adequate models while also suggesting that the model search should continue.” [32, p. 225]

“relativeness.” Explained in detail in Section 2.5 and Appendix A, relativeness emerged as the idea behind two of the eight (differently weighted) criteria used by the author to determine the clave direction of a vector as well as the degree of belonging to said clave-direction category. Patterns that exhibit this characteristic are highly representative²³ borderline cases. That is, they are *clearly* in a given clave direction, yet very close to being neutral. Hence, they are perfectly suited to determining where a cutoff ought to be for neural-network output values in terms of accepting or rejecting membership in an output category.

Experiments were carried out training a 5-hidden-element MLP with various training sets, including all 10784 vectors. (It can be argued that this does not violate Hastie, Tibshirani and Friedman’s warning because it applies equally to all designs, and because it is not a design step, but part of understanding the problem domain. Furthermore, this step was necessary because of the bootstrap/looping nature of all neural-net experimentation.)

One problem with this determination is that to do it with perfect accuracy would require all the other experimentation of this research to have been carried out first. Then, the resulting best network would be used to find the best threshold. However, the threshold value is required for all other experimentation to be done in the first place. This is the essence of the trouble with Machine Learning: All parameters need to be known before any of the parameters can be known optimally. Fortunately, the success of Machine Learning methods in solving problems to a practically sufficient

²³ of one clave-direction or another

degree means that optimality (as pursued in Statistical Learning) is traded off for practicality²⁴. Hence, a practical decision was made to use one particular MLP design to try out some training and test cases, and a reasonable threshold was found: 60%.

In one of the experiments where training included *all* available vectors, the following pattern rated an output of 60.0016%.

0101 1101 0010 0010

This pattern exhibits relativeness between the first and second, and between the first and fourth quarters of the phrase, as well as cancellation between the first and second, and between the third and fourth. Hence, not only is it an excellent example of relativeness in clave, it is also a borderline case because its clave direction comes down to one single onset (the downbeat of 2) and one single schema: 0101. These indications are in agreement, and the pattern is clearly in the FWD class (category 1).

However, the pattern scoring the closest output to this one, at a level of 57.8188%, was 1100 1111 0010 1011. This pattern only contains one decisive schema (0010) that leans it toward category 1, but the eighth criterion (“hanging,” supplied by the *a* of 2) suggests category 2. Cancellation between quarters 2 and 4 would also indicate category 2, but the weak schema of 1100 at the start pushes for category 1. From such an analysis, we can see that this pattern is not nearly as obvious in its clave direction.

²⁴ This is part of the culture of Machine Learning, and the main reason for the “culture wars” between the more mathematically leaning Statistical Learning academic culture (which seeks provable optimality) and the more engineering-oriented Machine Learning academic culture which seeks to find the best practical solution available under the limitations and resources available. This very “culture war” was played out in the late stages of proposing the present research: A committee member left because the proposed research was to pursue a good solution, not the optimal solution.

The next output, at about 56%, belonged to a neutral (category-3) pattern. Hence, the decision was made to set the threshold for all experiments to 60%.

In addition to concerns about the technical details of neural nets and Reconstructability Analysis, there are also issues of scope in regards to the acquisition of the data. The determination of proper constraints based on musical context is central to all forms of computer recognition of music or musical information. Musical signals are simply too complex in terms of the information they carry to approach the problem otherwise. The data for the proposed Neural Networks research is drawn from a music-recognition task. The classification of vectors to be used in the development of NN-design methodology (prestructuring) come from the music-recognition task carried out by the domain experts. Such a music-recognition task is best carried out with limited technical, cultural, and artistic scope because of the aural and idiomatic complexity of music signals.

Hence, there are a number of constraints common to the music-recognition literature. As evidenced in the literature review in the proposal for this research (provided in Appendix C), it is necessary and standard in technology development for automated music recognition to employ severe constraints on data generation. These have ranged from using only MIDI streams to limiting the genres and cultural domains allowed, to only considering monophonic inputs or solo instruments, to limiting the tempo range. Usually, more than one such constraint is used at a time in order to project multi-faceted musical data down to a manageably small set of attributes for the desired abstraction.

The data constraints that were necessary for the present research were similar in nature and number. In order to generate the data vectors, genres and styles under consideration were limited to those falling within the Afro-Brazilian folkloric idiom of *samba carioca*²⁵. Both in terms of the theory of musical grammar put forth as part of this research and in terms of the data preparation, time signatures were limited to duple; triple and other compound meters and triplet subdivisions in duple meters were ignored. Future work in the musicological aspect of this research can include expanding past these limitations.

1.16 RA-Search Heuristics

Just as the Neural Networks and Ethnomusicology aspects have necessary constraints, the use of Reconstructability Analysis in the present research suffered from limitations of today's computing power. Since the number of variables used to encode timing/accents patterns and their clave-direction classes have varied between 17 and 20, inclusive, a complete search (with even one RA criterion) is not possible with today's computing power. However, through heuristics employed in the RA field, smaller but still powerful searches can be executed.

²⁵ The author's original ambition was to develop not just a grammar of Rio-style samba, but one of all clave-based and African music in the world, from the Middle East to southern and central Africa (as well as West Africa, the source of clave), to Trinidad, Jamaica, Cuba, Belize, Uruguay, New Orleans, and even African-American funk. While the author's investigation into this possibility continues, some problems identified thus far (mainly in reconciling the current theory with evidence from the music of Uruguay, Cuba and New Orleans) would have held back the engineering aspect of the research. Hence, the author is grateful to Dr. Hansen for recommending that the cultural scope be highly narrowed. This was done so as to first reduce the scope to Latin America, then to Brazil, then to Rio de Janeiro, and then to samba as codified between 1888 and 1938, and practiced as "traditional" up to this day [47].

In addition, it would appear that, aside from computational limitations, the algorithms of Reconstructability Analysis themselves, perhaps due to the very nature of statistical significance and hypothesis testing (see Section 4.5), present a limitation in the models supply by *Occam3*. The 16-bit rhythm vectors that form the input portion of the I/O data have quarterly (90°) symmetry. Yet many of the models investigated and recommended by RA procedures have been “top heavy” in the variables. Considering that the entire length of the phrase must be considered even by a Brazilian expert (not because of the time it takes to decide or “feel” the clave direction, but because considering the whole is essential to the operation of clave direction), it seems unlikely that models like the following would embody clave-direction relationships: IV:ACZ:AJKZ:CGZ:FGZ:KMZ:QZ:RZ²⁶, which shows higher-order interrelationships among the earlier attack positions in the rhythm cycle (such as A, B, and C, rather than P, Q, and R), and IV:ABZ:ACZ:AKZ:ASZ:BDZ:BQZ:BRZ:CDZ:EFGZ: FJZ: HJZ:JKZ: JNZ:JQZ:JRZ:KMZ:PZ:QSZ:RSZ, which exhibits more numerous higher-order interrelations at the top end than in the neighborhood of P, Q, R, and S.

Instead, statistical heuristics have called off the search at this point, before a model with an evenly distributed selection of interacting components could be found.

One model that did not have this problem, IV:ABCQZ:ABRZ:BFJQZ: BMQRZ:CJKZ:CKRZ:FGZ:GRZ:JKMZ:JKQZ:JKRZ: JQRZ:KMQZ:KQRZ, where

²⁶ For the assignment of these variables, see Section 4.2. Also, to be precise, in RA terminology, such a string alone is a *structure*, not a model. A model is *structure* with *data*. Since all structures referred to in this dissertation are intended for use with the research data, this distinction is not followed in the rest of the document for ease of transferability of ideas and statements between the RA realm and the model-selection/model-evaluation literature.

the three- and four-way interactions of input variables are more evenly distributed than the aforementioned models, presented a whole other problem: It has 65473 degrees of freedom, which is almost exactly equivalent to having a look-up table for each of the 65536 possible input vectors.

Practical RA searches (if including all model types) require the specification of search parameters called *width* and *level*. In searching through the tree of models, whose branches number in the millions for a 17-variable system, width specifies how many of the best-predicting models at a given structural level *Occam3*²⁷ will keep in the search and how many it will discard. For example, for a width of 5, the five top models at each level of the tree will be kept, and the candidates from the next level of the tree will only be generated from the ancestors or offspring of those five models, and so on. Similarly, level gives *Occam3* an upper bound on how far up or down the tree of models to travel during a search.

In trial runs, individual searches of high width and level (depth) have taken 24–72 hours to run. In order to reduce the risk of not finding a large portion of good models without increasing computational load to a point that would dominate the *Occam3* server for long periods of time (partly out of concern for other users), various combinations of search types, widths and depths can be strategically combined. There are two basic strategies. These are the *loopleess bore–rotate search* and the *all-model bore–expand search*.

²⁷ Version 3 of the Organizational Complexity Computation And Modeling software developed and maintained at the Department of Systems Science, Portland State University.

When using feature selection, the loopless bore–rotate search starts with a loopless search for feature selection before moving to an all-model search for structure discovery:

Vertical Loopless Search: Set width to 1 or 2; set to the number of variables.

Use loopless bottom-up search to decide how many levels up to search.

Rectangular Loopless Search: Set the level to the optimal found in the previous step. Set the width very high (around 20). Perform another loopless bottom-up search.

Rectangular All-Model Search: Turn off all predictively unimportant variables and perform a bottom-up all-model search.

When not using feature selection, the same steps are followed with the exception of turning off predictively unimportant variables. This distinction is a major cause for concern because on one hand, the expert-recommended RA procedure is to remove apparently insignificant input variables, but on the other hand, the present author-as-clave-expert is convinced that all input variables take part in determining clave direction, even if some matter less than others.

This is a case of Occam's Razor as expressed in its technological form by the Lendaris/Stamley conjecture in terms of overfitting. Even though it is true that all variables matter, if some matter more than others, better generalization is likely to be achieved when not paying attention to every detail of the training set. This is the classic example of overfitting in regression: The curve that matches every point in the plot of training examples is certain to overtrain for generalization to unseen cases. Hence, the

preference for the primary research direction is to include feature selection when suggested by RA search results.

The all-model bore–expand search uses a width-1, all-level, all-model search to determine the point after which further search levels do not result in improved models. This is then followed by searches up (or down) to the level thus determined, but with increasingly high widths.

Furthermore, it is possible to start *Occam3* searches at any specified point in the middle of the tree of models and proceed in either direction. The incremental- α feature is specifically helpful with this technique, allowing the significance of individual steps to be judged with respect to neighboring models, whereas cumulative α compares all models considered to the reference.

1.17 Definitions

1.17.1 Multilayer Perceptron (MLP)

The quintessential artificial neural network, MLP, is among the most commonly employed neural-net types because it is applicable to any problem in which the outputs depend on many inputs, and the inputs and outputs (in a sense, questions and answers) can be stated in numerical form. Fashioned after a somewhat simplistic version of the neuronal connections in small portions of the brain, a trained MLP is not a general-purpose thinking engine like the brain, but a specialized problem-solver. However, much like the brain which can redeploy its functional units for completely new tasks (motor, emotional, or other), the MLP can also be retrained for different tasks.

The plasticity of the brain is essential for such redeployment to pay off, and methods (beyond the regular training modification of changing the connection weights linking existing neurons) have been devised to grow and prune artificial neural nets.

Technical detail about the MLP is in Section 2.3.

The prestructuring of MLPs has been shown to be effective in improving their generalization performance on artificial data sets. The goal of the present research is to expand the scope of prestructuring-for-generalization-improvement to real-world problems.

1.17.2 Generalization

In Computational Intelligence (CI), generalization is the process of induction applied to new inputs, questions or queries. A supervised CI agent is trained with examples of queries that, if well chosen, represent the full variation of facts and circumstances in the system the agent is trained to learn. Based on the (mathematical) learning procedure the agent follows, aggregating the input/output relationships discerned in the training data, connections in the agent's internal network are formed and reinforced so as to reflect the structure of the problem or system being learned—to the extent that the available examples reflect the features of the problem or system.

After this training process, the CI agent is presented with data (input, questions, or queries) that it had not encountered in training. The outputs (responses or answers) of the agent are its generalization. The agent may be right or wrong. The more often an agent is right in responding to inputs it had not encountered in training, the higher its generalization *performance* is.

Generalization performance is discussed in detail in Section 2.1.

1.17.3 Clave Direction

Clave direction is a rhythmic property of a musical phrase that fulfills certain criteria. For example, while clave direction may also exist in odd-metered phrases, this has not been established at any cultural or scholarly level. On the other hand, a verifiable lack of clave direction may indeed be a type of clave direction because it is aesthetically desirable (and perceptually helpful) for clave-based music to contain many clave-neutral elements for rhythmically grounding the listener (or performer or dancer) and providing contrast and movement for the parts of the music that do exhibit clave direction.

The reader has likely noted that a definition of clave direction has not yet been given. This is because clave direction cannot be defined quickly and simply. The typical efforts in this regard have mostly served to confuse the issue and mislead students of clave: “There are three beats on one side and two on the other.”

But what exactly is a beat? And more importantly, what happens when there is one on each side, or when there are five on one side and seven on the other? Or when there are three on the so-called “2-side” and two on the “3-side”? This last case can be heard in some examples of Belizean music, as shown in Figure 9, and in Brazilian music, as in Section A.2.6.4.

Clave direction is a multi-faceted concept that has been subject to revision over time and space; i.e., throughout history and around the world.

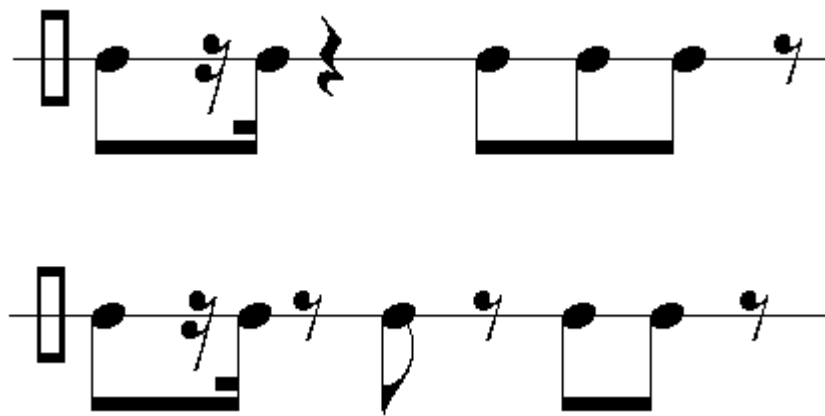


Figure 9: Belizean “3-2” clave (top) and the common *son* clave (bottom) in the same “direction.”

The version of clave direction used in this dissertation is derived from detailed study of the Afro-Brazilian samba tradition of Rio de Janeiro, specifically as it evolved following Brazil’s abolition of slavery in 1888, through the subsequent *legitimization* (legalization, and then nationalization) of samba in the late 1930s [49, including a discussion of how Angolan samba may have acquired its Yorùbá-derived sense of clave direction, pp. 88–9²⁸].

For details on clave direction, see Section 2.5 and Appendix A.

1.17.4 Tactus and Tatum

The tatum, introduced in [50], has to do with the smallest timing interval found in a piece of music. It is complicated by tempo variations, expressive timing, rolls and ruffs, and possibly other musically expressive actions. To understand tatum, one can

²⁸ A summary of the hypotheses regarding the binarization of ternary African rhythms is given in [51].

start considering fully quantized computerized music with no tempo changes or variations, and build toward more complex conditions.

Imagine a piece with two instruments whose timing is represented below irrespective of pitch (Figure 10).



Figure 10: Two parts (to be played simultaneously) for demonstrating tatum-finding.

The resultant 16th-note tatum from these two simultaneous rhythms (vertically collapsing the two and assuming no other interaction in the rest of the piece of music) is indicated by Figure 11.

The tactus, on the other hand, is the most typically felt pulse in a piece of music: the beat that we tap our feet to, which corresponds to the lower number in the time signature. The tactus may be divided up into tatums if the rhythm is metronomically steady.



Figure 11: Even though there are no 16th-note interonset intervals (IOIs, [52, p. 4]) on either instrument, the tatum for the overall piece is of 16th-note duration because of the interaction between the two parts in the second bar.

The term *tatum* is now the academic standard for Kwabena Nketia's "density referent" [53, p. 44].

1.17.5 Syncopation and Offbeatness

Coined by Toussaint as a measurable quantity, the concept of offbeatness [54] has proven invaluable to the present author's preliminary work of understanding Afro-Brazilian (*partido-alto*) clave direction. It is interpreted here as a more precise term for rhythmic purposes than the term syncopation would be. (Syncopation has a formal definition that is culturally rooted: the placement of accents on *normally*²⁹ unaccented notes, *or* the lack of accent on *normally* accented notes.) It would be reasonable to assume that the norm in question is that of the genre, style or cultural/national origin of the music under consideration. On the contrary, one finds that in its usage around the world (except the present author's), "normal accent placement" is overwhelmingly taken to be normal *European* accent placement [12, 55, 56].

For example, according to Kauffman [55, p. 394], syncopation "implies a deviation from the norm of regularly spaced accents or beats." Various definitions by leading sources cited by Novotney also involve the concepts of "normal position" and "normally weak beat" [12, pp. 104, 108].

Thus, syncopation is seen to be norm-referenced, whereas offbeatness is not contextual—it depends solely on the *tactus*.

²⁹ usually, typically, commonly in the idiom.

Kerman, too, posits that syncopation involves:

... accents in a foreground *rhythm* away from their normal places in the background *meter*. This is called **syncopation**. For example, the accents in duple meter can be displaced so that the accents go on one *two*, one *two*, one *two* instead of the normal *one two*, *one two* [56, p. 20; all emphasis exactly as in the original].

Similarly, on p. 18, Kerman reinforces that “[t]he natural way to beat time is to alternate accented (“strong”) and unaccented (“weak”) beats in a simple pattern such as *one two*, *one two*, *one two* or *one two three*, *one two three*, *one two three*.” [56, p. 18]

Hence, placing a greater accent on the second rather than on the first quarter note of a bar may be sufficient to invoke the notion of syncopation. By this definition, the polka is syncopated, and since it is considered the epitome of “straight rhythm” to many performers of Afro-Brazilian music, syncopation clearly is not the correct term for what the concept of clave direction is concerned with. Offbeatness avoids all such cultural referencing because it is defined solely with respect to a pulse, regardless of cultural norms. (Granted, what a pulse is may also be culturally defined, but there is a point at which caveat upon caveat becomes counterproductive.)

Furthermore, in jazz, samba, and reggae (to name just a few examples) this would not qualify as syncopation (in the sense of accents in abnormal or unusual places) because beats other than “the one” are regularly accented in those genres as a matter of course. In the case of folkloric samba, even the placement of accents on the second

eighth note, therefore, is not syncopation because at certain places in the rhythmic cycle, that is the *normal*—expected—pattern of accents for samba, part of the definition of the style. Hence, it does not constitute syncopation if we are to accept the definition of the term as used and cited by Kauffman, Kerman, and Novotney. In other words, “syncopation” is not necessarily the correct term for the phenomenon of accents off the downbeat when it comes to non-European music.

Moreover, in *Meter in Music*, Hule observes that “[a]ccent, defined as dynamic stress by seventeenth- and eighteenth-century writers, was one of the means of enhancing the perception of meter, *but it became predominant only in the last half of the eighteenth century* [emphasis added]. The idea that the measure is a pattern of accents is so widely held today that it is difficult to imagine notation that looks modern and that does not imply regular accentual patterns. Quite a number of serious scholarly studies of this music [European art music of 1600–1800] make this assumption almost unconsciously by translating the (sometimes difficult) early descriptions of meter into equivalent descriptions of the modern accentual measure” [57, p. viii] Thus, it turns out that the current view of rhythm and meter is not natural, or even traditional, let alone global. In fact, in *Essential Dictionary of MUSIC NOTATION: The most practical and concise source for music notation is perfect for all musicians—amateur to professional* (the actual book title) states that “the preferred/recommended beaming for the 9/8 compound meter is given as three groups of three eighth notes” [58, p. 73]. This goes against the accent pattern implied by the 9/8 meter in Turkish (and other Balkan) music, which is executed as 4+5, 5+4, 2+2+2+3, etc., but rarely 3+3+3. The 9/8 is one of the most common and

typical meters in Turkish music, not an atypical curiosity. This example and the related reference are included here to demonstrate the dangers in applying western European norms to other musics (as indicated by the phrase “*perfect for all musicians*”).

1.17.6 Attack-Point Rhythm

This concept is most clearly explained through the words of a music teacher upon seeing the 16-bit representation of attack-point rhythm: “Each grouping of 4 as a beat, and the 1 representing a note (not necessarily a 16th, just a note appearing at this moment. Could be a dotted eighth if followed by three zeros)³⁰” [59]. A selection of the patterns this teacher was looking at is given in Figure 12.

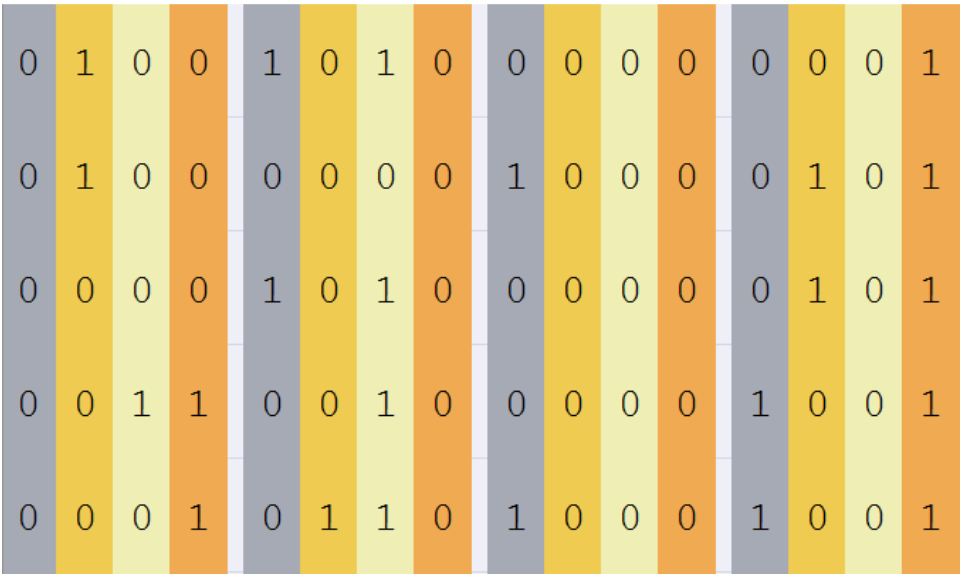


Figure 12: A few of the attack-point patterns presented to participants in the human-benchmark studies; the shading and large font size are for participants’ ease of reading. Support for this representation is found in industry [60], academia [11], and among professional musicians [61].

³⁰ Correction: A 1 followed by two 0s and another 1 implies a dotted 8th note with more certainty.

CHAPTER II. BACKGROUND

This chapter is a review of background information on Computational Intelligence, Information Theory and clave direction, which are the primary components of the present research.

Section 2.1 places the research that was undertaken in its scientific, technological and cultural context, and to provide a structural interpretation of the field of Computational Intelligence within which it is situated.

Section 2.2 breaks down the technique of cross-validation (central to neural-net design) into its types and applications, and also includes an explanation of the primary performance metric proposed for neural networks.

The cross-validation technique is at the philosophical heart of this dissertation because one of the key questions raised by the results is whether cross-validation is a default aspect of neural-net design or whether it ought to be considered a means of learning from the data or the problem domain.

Section 2.3 provides the theoretical and historical background on Neural Networks, which is the primary topic of the present research. This background is necessary in order for one to make sense of both the literature search and the experimental process and results.

Section 2.4 is an explanation of Reconstructability Analysis (RA), which combines multiple ways of modeling relationships among variables, and implements various trade-offs between accuracy and generalizability, as reflected in the information-

theoretic quantities of Constraint and Entropy. (The reasons for capitalization are given in the relevant sections.)

Section 2.5 goes into a more detail on clave direction (though a yet more involved treatment is in Appendix A) than given in the definitions in Chapter I, and Section 2.7 ties the four main theories underlying the present research into a cohesive whole: Artificial Neural Networks, Information Theory, Reconstructability Analysis, and clave direction.

Finally, Section 2.7 presents a brief discussion of the author's credentials for doing the research presented here, and is included to place the research in the context of the author's history of teaching and learning, with which it is intertwined.

2.1 Science and Computational Intelligence

In his chapter on empirical research, Cohen identifies three types of research questions and four types of empirical studies, placing them within a space of scientific progress whose coordinate axes are *understanding* and *generalization* (Figure 13) [62, pp 4–5]. The three types of research questions are [62, p. 4]:

How will a change in the agent's structure affect its behavior given a task and an environment?

How will a change in an agent's task affect its behavior in a particular environment?

How will a change in an agent's environment affect its behavior on a particular task?

What follows are examples of each type of research question, one from an unrelated but generally familiar domain, followed by one for the present research.

1a. How will a change in one constituent of an antibiotic alter its effectiveness to cure a given condition in adult males?

1b. How will disallowing certain connections among processing elements affect the generalization performance (behavior) of a multilayer perceptron given a task of recognizing Afro-Brazilian, *partido-alto* clave-direction when trained under a particular set of assumptions (environment, here called “the teacher model”)?

2a. How will the original (unaltered) antibiotic perform in curing a different condition in adult males?

2b. How will replacing the training and test data with those of the classic iris set affect the generalization performance of a prestructured neural net designed for the clave problem?

3a. How will the original antibiotic perform in curing the original condition in adult females?

3b. How will replacing the training and test data for the firm-teacher model with those relating to the strict-teacher model affect the generalization performance of a network prestructured according to models of the firm-teacher data on the new data?

Each of these types of research questions can be investigated through one of four types of empirical studies [62, p. 7]:

“Exploratory studies” occur in the early stages of the scientific method³¹, and “yield causal hypotheses that are tested in observation or manipulation experiments.”

“Assessment studies” also occur in the early stages of research, and serve to “establish baselines and ranges,” (such as the ‘floor’ and ‘ceiling’ determinations with uninformed and informed human agents—respectively—undergoing the training and test procedures of the present clave-direction-recognition problem).

“Manipulation experiments” involve the types of statistical hypothesis testing that are typically associated with proper science by manipulating one factor at a time and observing the changes in dependent variables.

“Observation experiments” look for correlations between the manipulation of factors and the variations in dependent variables. They are used to generate or assess models. Hypothesis testing is common but not a requirement. (Parts of all four explanations are quoted or paraphrased from [62, p. 7].)

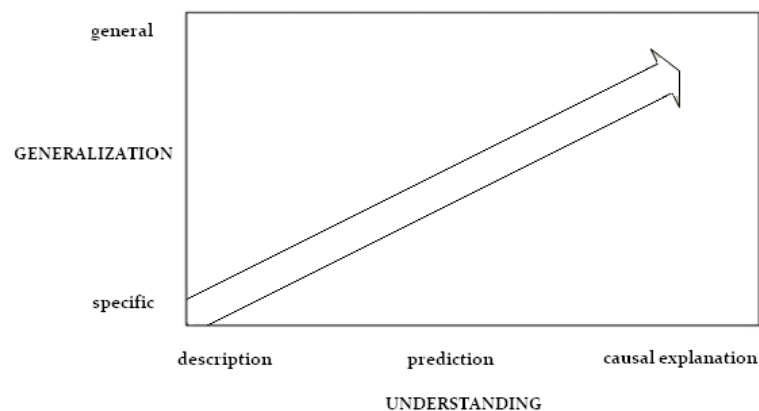


Figure 13: Cohen's diagram of scientific progress (recreated).

³¹ See Chapter III: Literature Review, and Chapter IV: Methodology for detailed discussions of statistical techniques and tools relevant to the experimental design of the present research.

The exploratory studies that set the stage for this dissertation come from three sources: the prestructuring of artificial data (cf. Lendaris [63–67]), the application of Reconstructability Analysis to prestructuring [30, 48, 64, 68], and the author’s initial studies of clave in general (since 1992) and then Afro-Brazilian traditional samba in particular (since 1997), as described in Appendix E and Appendix I.

Assessment studies were conducted with two groups of human agents. The ‘ceiling’ group consisted of paid mid-level experts in *partido-alto* clave direction who are professional or semi-professional performers and teachers of Afro-Brazilian music in the United States. Only four such agents were available out of a total of seven in the state of Oregon, to the researcher’s knowledge. One is the researcher himself (thus ineligible), and two were unavailable. The remaining four took part in the study, and one of those four dropped out partway through.

The ‘floor’ group consisted of a large initial pool uninformed persons³², either nonmusicians or non-samba musicians, undergoing the same training and test regimen as the ‘ceiling’ group, who received a proportionate, representative sampling from the training and test sets for the neural-net portion of the research³³. This was a self-selected convenience sample, which is deemed appropriate because all uninformed persons—whether self-selected or not—are expected to perform at roughly the same generalization level for such an abstract task as classifying patterns of zeros and ones

³² This large group dwindled to a small number who actually completed the task.

³³ The reason for the proportionate sampling is that people who are not fully invested in this research are typically not willing to look at thousands of strings of 1s and 0s.

based on examples, especially in the absence of information about what those patterns represent.

In terms of time spent, the bulk of the present research has been defining and understanding the problem and the techniques (RA, statistical rigor, NeuralWorks, and *Occam3*), and the subsequent manipulation experiments. Factors manipulated included the output representation, the input representation (in terms of feature reduction via *Occam3*), the number of network hidden-layer elements, various neural-net initialization and regularization parameters, input-classification models, and training/test splits. Some three years from the acceptance of the dissertation proposal to the writing of the final dissertation were spent understanding the problem by alternately manipulating these factors, and investigating the results of various NN and RA runs. These manipulations were informed, in order of significance, by the NN experience of dissertation adviser Dr. Lendaris, the domain insight of the author, the NeuralWare reference materials provided with NeuralWorks, the extensive Neural Networks reference and textbook by Haykin [26], several other leading books on experimentation, statistical learning and neural networks [27, 39, 69] and an extensive collection of Statistics and Machine Learning literature from peer-reviewed journals and conferences. These are acknowledged throughout this dissertation as their contributions are cited.

A discussion and examination of the theoretical bases and practical tools of the field of Statistics is given in Chapter III (Literature Review) and Chapter IV (Methodology).

2.2 An Overview of Machine Learning and Statistical Learning

This dissertation concerns the field of Artificial Neural Networks, or simply Neural Networks (NN), which is a subfield of Computational Intelligence (CI), which can be considered a further subfield of Machine Learning (ML), which in turn forms one of the two principal branches (Figure 14) in the research and application of high technology to complex problems of automated search, optimization (design), forecasting (prediction) and classification (understanding). This section gives a brief overview of the structure of the overall domain of research, which combines Computer Engineering, Computer Science and Systems Science through the use of tools and inspiration from Statistics, Mathematics, Physics (thermodynamics), Biology (zoology, evolution and genetics), meme theory³⁴, cognitive and experimental Psychology, and Philosophy (formal logic and philosophy of science).

From modeling naturally complex economic, geological, meteorological and biological systems, to creating complex virtual environments, adaptable learning machines, and unmanned vehicles, the STEM³⁵ fields have been engaged in the pursuit of machine intelligence (however approximate or philosophically debatable) for several decades. In this period, modern technology³⁶ has expanded its range of functions from purely logical and systematic applications like numerical computation and word-processing, into the modeling, mimicking, and even design of highly complex, non-

³⁴ Meme theory is likely not an actual science, but its technological usefulness overrides its possible lack of scientific validity, just as many algorithms in the field of Evolution Strategies use physically impossible or biologically unrealistic tricks to obtain technologically useful solutions.

³⁵ A common acronym for “science, technology, engineering and mathematics” used in the education field and the federal government.

³⁶ For the sake of argument, let’s take “modern technology” to begin with the transistor.

deterministic systems. The latter range from the early applications of adaptive signal processing to echo-cancellation to mainstream speech-recognition techniques to the ongoing development of information retrieval through the semantic web, and error-tolerant computing (also known as stochastic processing [70]).

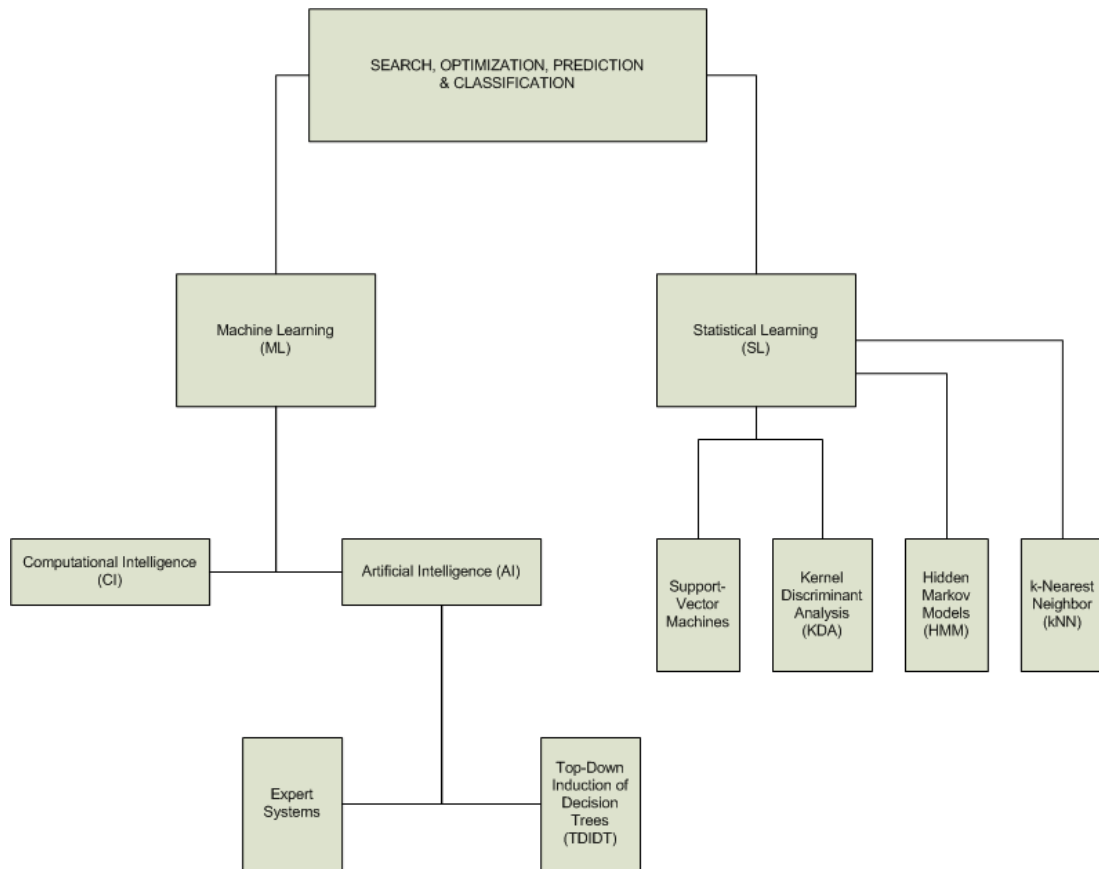


Figure 14: A possible partial classification of engineering approaches to search, optimization, prediction (forecasting) and classification (pattern recognition) problems. These divisions, even between Machine Learning and Statistical Learning, are not absolute—there is overlap among many of these categories, both in terms of methods and in terms of the underlying philosophies.

The main fields in this endeavor can³⁷ be broken up into Machine Learning (ML, which consists of Computational Intelligence and Artificial Intelligence) and Statistical Learning (SL)³⁸, which includes Kernel Discriminant Analysis (KDA), Hidden Markov Models(HMM), k-Nearest Neighbor (kNN) and Support-Vector Machines (SVM).

All studies of complex systems rely on induction. Machine Learning and Statistical Learning differ in their mode of induction. Statistical Learning is thoroughly founded in theory; Machine Learning is primarily empirical. This difference shows up in the cultures of the two fields in that practical results that are not solely based on theoretical foundations are not *systemically* rejected in Machine Learning, where “getting the job done” supersedes guaranteeing optimality.³⁹

However, with all their differences, the two fields are intertwined both in terms of their applications and in terms of being of service to one another. For example, Statistical Learning provides many theoretical bounds and parameter-selection guidelines that can inform experimental design in Machine Learning. Indeed, many such guidelines and results from Statistical Learning were used in the present research to steer and delineate the choice of design and comparison procedures.

The two types of Machine Learning, namely Computational Intelligence (CI) and Artificial Intelligence (AI), differ in their design approaches. CI uses a biologically/

³⁷ “Can be” because these distinctions are neither precise nor universal. These fields have varying degrees of overlap with each other in terms of methods and philosophy.

³⁸ SL is further divided into Parametric and Non-Parametric Estimation, the details of which are beyond the scope of this document.

³⁹ For more on the differences and similarities as well as the strengths and weaknesses of each approach, see [172].

psychologically/socially inspired bottom-up approach; AI uses a technologically inspired top-down approach⁴⁰.

The bottom-up, biologically inspired CI subfields include Artificial Neural Networks, Evolutionary Algorithms (which are of five primary types, augmented by numerous hybrids, co-evolutionary algorithms and social/memetic algorithms), the related fields of Fuzzy Logic, Rough Sets and Computing With Words (CWW), Cellular Automata and Artificial Life⁴¹, Immune-System Algorithms, Swarm Intelligence⁴² (successfully used for modeling and problem-solving in telecom routing and public transportation), and others.

The present work primarily investigates a further subfield of CI: Artificial Neural Networks, or simply Neural Networks (NN).

The standard method of model selection and model assessment in NN, and in fact throughout Machine Learning and Statistical Learning, is cross-validation [26, pp. 213–8; 29]. Cross-validation means the use of some out-of-sample test set for evaluating the true (future) performance of a hypothesis (a network, a decision tree, a discriminant, etc.)⁴³. The out-of-sample test set is held separate from all training and design steps, and hence also called the *holdout*. This type of assessment (estimation of generalization

⁴⁰ The older top-down models, like Expert Systems, have given way to the current sense of Machine Learning, which is primarily associated with Top-Down Induction of Decision Trees (TDIDT), Bayesian Learning and Bayesian Belief Networks.

⁴¹ The difference between artificial life [71, pp. 69–71] and evolutionary algorithms [71, pp. 58–68, 71–72] is explained with a clear comparison in Dawkins' *Climbing Mount Improbable* [71].

⁴² Based on the behaviors of ants, bees, termites, birds and fish.

⁴³ 'Hypothesis' is the scholarly equivalent of the business term 'solution'. It refers to any candidate solution or design for a particular problem.

performance, in the case of Artificial Neural Networks) is essential to the scientific validity of any CI, ML or (nonlinear) SL study⁴⁴.

2.2.1 Types of Cross-Validation

There are two variants of cross-validation, known as *Holdout Cross-Validation* and *Multifold (k -Fold) Cross-Validation*. The following explanation of these techniques and the delineation of their corresponding data splits is a synthesis of material from Haykin's *Neural Networks* [26], Hastie, Tibshirani & Friedman's *The Elements of Statistical Learning* [39], class notes and discussions from PSU's CS 410/510 *TOP: Machine Learning*, 2005 (Mitchell) [72] and OGI's (*Advanced Topics in*) *Machine Learning*, PSU CS 410/510/OGI CSEE 559/659, 2005 (Leen) [29], and aided by discussions with Dr. Lendaris, the author's adviser.

As shown in Table 1, the two types of cross-validation can each be used for either model selection or model assessment.

Model selection constitutes the bulk of NN design: In the course of a series of experiments, multiple networks with different numbers of hidden-layer elements, regularization parameters, and other attributes are developed according to heuristics developed in the fields of NN and SL. One of these candidate networks is chosen as the final design. Model assessment is the determination of a performance value that

⁴⁴ "In regression models *linear in the parameters*, there are *algebraic estimates* of generalization error given the training error. In models *nonlinear in the parameters*, this is far more difficult[,] and we turn to *empirical methods*." [30, slide 7].

estimates the final design’s future performance according to the state of the art in estimation science, i.e., Statistics.

Table 1: Types and Applications of Cross-Validation

	Cross-Validation Technique	
Application	k -Fold for Model Selection	Holdout for Model Selection
	k -Fold for Model Assessment	Holdout for Model Assessment

The experimental design for the present research uses k -fold cross-validation for parameter estimation (designing the network, also called “model selection”) and a single holdout for performance estimation (evaluating the designed network, or “model assessment”).

The reasons for these choices are that k -fold cross-validation allows flexible and repeated use of a data set for the iterative and unpredictable process of designing the networks, and that a large, true holdout set put aside prior to any design work is the best possible means of performance estimation⁴⁵, which is the ultimate outcome of the present research. Since both methods are used, both are explained below.

⁴⁵ free from concerns about the choice of k or the size of the data set.

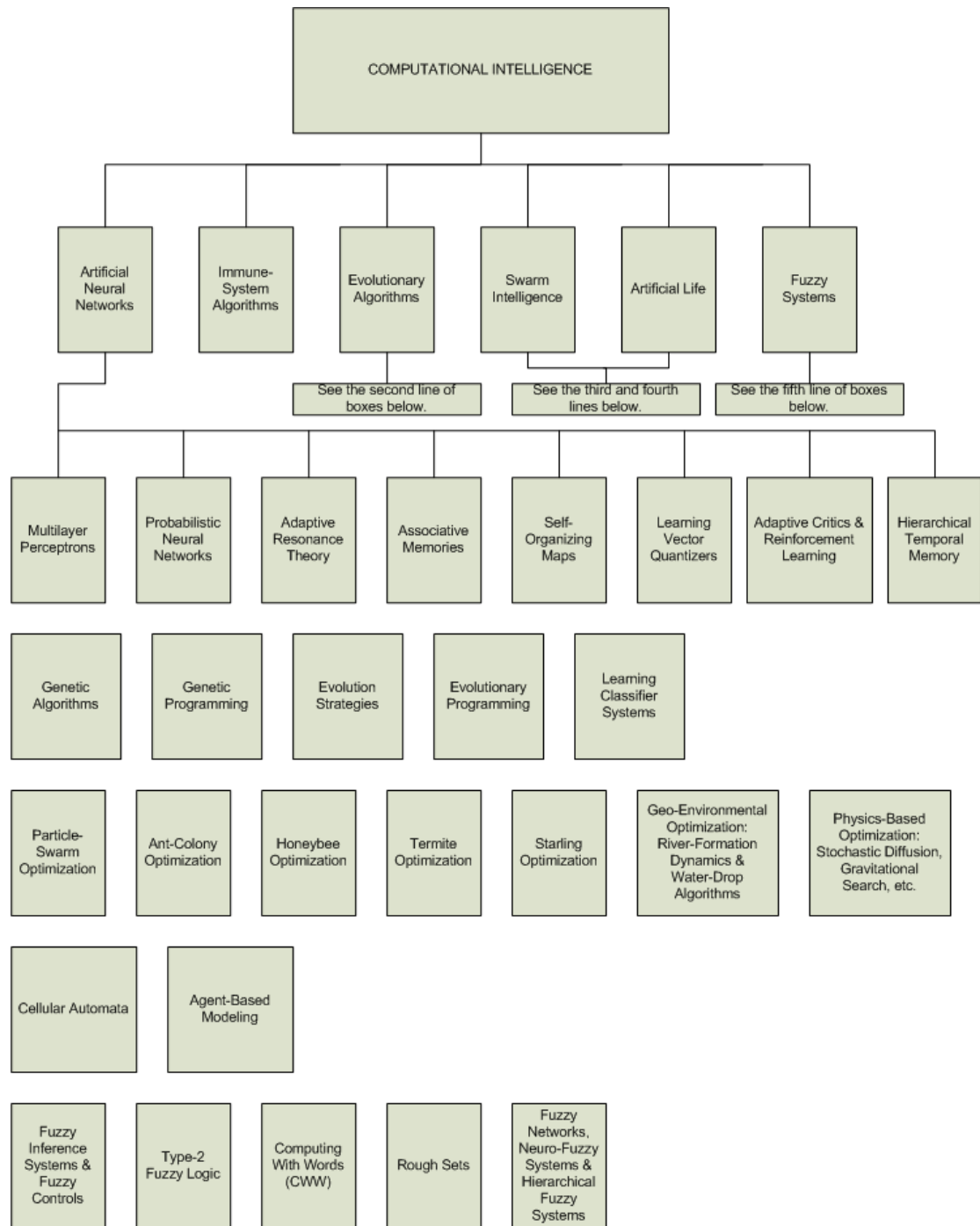


Figure 15: A partial family tree of subfields in Computational Intelligence. Not shown due to space constraints are the Generalized Regression Neural Network (GRNN), Counterprop, CMAC, the Infomax network, various types of Recurrent Neural Networks, and some Swarm algorithms.

2.2.2 Holdout Cross-Validation: The Basics

In Holdout Cross-Validation, the available data are divided into three subsets: the estimation subset, the validation subset, and the test set. The first two are collectively known as the development data set (also called “design data” in this document); the estimation subset is commonly called the training set; and the test set is alternately known as the holdout set or the out-of-sample test set.

2.2.3 Holdout Cross-Validation for Design

If this method is used for design (model selection), the network is originally trained using the estimation subset and tested on the validation subset for various design changes (network size, regularization parameters, early-stopping, or other parameters as discussed below). Once the design is finalized (“frozen”), the network is retrained with all the design data (estimation subset and validation subset) and its performance estimate is given by the performance of a test pass on the out-of-sample test set. The crucial feature of the holdout method is the holdout set whose elements do not take part in any design step.

2.2.4 k -Fold Cross-Validation: The Basics

In k -Fold Cross-Validation, there is still a design set and a test set, but the design set is divided into k equal subsets. Each of the k subsets (folds) is set aside as a temporary validation set while training is conducted using the other $k-1$ subsets. Thus, k separate trainings take place, with significant overlap in training data but no overlap in validation data. The performance estimate is given from the average of the k validation

results. As in Holdout Cross-Validation, an out-of-sample test set⁴⁶ is set aside before any supervised pre-processing (feature selection) is carried out on the data.⁴⁷ In this context, “supervised” means pre-processing is done in the presence of, or using, category labels [39, pp. 245–247], e.g., information-theoretic modeling with *Occam3*.

2.2.5 *k*-Fold Cross-Validation for Design

In the variant of *k*-Fold Cross-Validation for parameter estimation, design choices are made by averaging the validation results from each of the *k* folds. When the design is frozen, the entire (“unfolded”) design data is used to train the final network, and the generalization performance on test set is reported as the performance estimate for the network.

2.2.6 Statistical Soundness

In order to avoid the “blunder committed many times in published papers in top[-]rank journals” [39, p. 247], the RA searches for this dissertation research were carried out with the “meta-holdout” set taken out prior to the *Occam3* search/modeling operations.

⁴⁶ In the context of *k*-Fold Cross-Validation, the holdout set for the final performance assessment is referred to here as the “meta-holdout” in order to clearly indicate that it is external to *all* experimental processes except the final assessment.

⁴⁷ An alternate performance estimate is to choose the model—or network, or fold—that is within one standard error from the best-performing model. [29, p. 216]

2.2.7 Holdout Cross-Validation for Assessment

When designing networks with, and reporting on the results from holdout cross-validation (*model selection*), we choose the network design by testing repeatedly on the validation data. Then we recombine the design data for a final training of the frozen design, and report its performance on the holdout set (*model assessment*).

2.2.8 k -Fold Cross-Validation for Assessment

The variant used for performance estimation is for an already designed (frozen) network, and a further test set is not needed. The average of the k validation-performance values gives the proper estimate of network performance.

2.2.9 Motivation for Cross-Validation

The reason for test sets and holdouts in general is the “optimism” [39, pp. 228–230] of any training-set classification performance. Methods such as AIC and BIC are said to estimate the “optimism” of the training-set performance and numerically add this to the error given by the training set [39, p. 230]. AIC and BIC are two of the four criteria used to rate, rank and select models in Reconstructability Analysis. “In contrast, cross-validation and bootstrap methods ... are direct estimates of the extra-sample error ... These general tools can be used with any loss function and with nonlinear, adaptive fitting techniques” [39, p. 230] such as neural networks.

2.2.10 The Overall Holdout

For the validity of the comparison of generalization estimates for the prestructured and fully connected networks (the comparison that is central to this dissertation), an additional holdout set is needed. These data are set aside prior to any training of neural nets, and are in addition to any holdout set discussed in the standard techniques of Holdout Cross-Validation and k-Fold Cross-Validation.

2.2.11 Performance Metric

Performance assessment will employ the Generalizing Ratio. The Generalized Generalization Ratio (GGR, developed below and in the proposal for this thesis) would allow networks designed using different-sized training and validation sets, and compared on different-sized test sets to be compared in terms of their generalization performances, but it was possible to test all network types on the same-size test, training and holdout sets, so the Lendaris-Stanley Generalizing Ratio (GR) suffices. In fact, the Dunnett ANOVA to be used at the statistical-analysis stage requires sample sizes to be equal, and since the GGR is an extension of the GR for differing sample sizes, the ANOVA requirement means the experiments must be set up so that the GGR reduces to the GR. The development of the two ratios is given below.

Lendaris and Stanley offer metrics (and a definition) for generalization in their seminal 1965 paper: “Given two functions, F and G , with the same independent variables, if for every n -tuple (x_1, x_2, \dots, x_n) for which F is defined, G is also defined and assigns the same value as F , then G is an *extension* of F ” and “any extension of a

function is a generalization of that function.” [22, pp. 461–2] This definition does not involve a sense of performance accuracy, so Lendaris and Stanley define several ratios to be used as metrics of generalization performance. The conventions and variable names used to derive these metrics [22] are:

- I_r – the set of all possible inputs
- D – a target relation that the relation R approaches as a system’s parameters vary (such as the training of a neural network)
- R – the relation between the inputs and outputs of a system
- R_1 – a relation, like R , but defined on a component, S_1 , of system S
- C – the care terms of a partial function⁴⁸ D
- C_1 – a proper subset of C (specifically, the training set)
- $C \setminus C_1 = \{x \mid x \in C \text{ and } x \notin C_1\}$
- c – cardinality of C
- c_1 – cardinality of C_1
- $k = c - c_1$
- k_i – the number of care terms in $C \setminus C_1$ for which R_1 agrees with D

Hence, k_i/k is an indicator of generalization quality in terms of the degree of correctness of R_1 ’s mapping of care terms in D that are not in the training set C_1 .

Generalization performance, then, can be quantified with the two-fold

expression: $\frac{k_i}{k}$ generalization via $\frac{c_1}{c}$ exposure.

According to this, while better generalization is always better, equal generalization under less exposure is superior [22, p. 462]. According to Lendaris and Stanley, in most cases “ c and k are very large numbers” [Ibid.] to the point of making

⁴⁸ A partial function is one that specifies a one–one (or many–one) mapping from a domain to a range for a subset (not necessarily proper) of the elements of the domain.

the above metric impractical. Instead a representative, smaller set E^{49} , the exam set, is selected from $C \setminus C_1$, its cardinality denoted by e . Analogous to the definitions above, e_c represents the care terms for which the system outputs the correct classification.

Now we can talk about $\frac{e_c}{e}$ generalization via $\frac{c_1}{c_1 + e}$ exposure for a practical approximation [22, p. 463]. To clarify that C_1 is the training set, let's call the number of correct classifications in the training set t_c with $t \equiv c_1$.

Hence we have “ $\frac{e_c}{e}$ generalization via $\frac{t_c}{t}$ training performance on $\frac{t}{t+e}$ exposure.” [Ibid.]

In order to be able to compare different networks (hypotheses, agents, etc.) where neither of the levels of generalization and exposure is (necessarily) held constant from one network to another, Lendaris and Stanley propose the “Generalizing Ratio” [Ibid.]:

$$GR = \frac{e_c / e}{t_c / t} \text{ for constant } \frac{t}{t+e}. \quad (1)$$

It is proposed here that the assumption of holding the training and exam set sizes constant can also be relaxed with the following “Generalized Generalization Ratio”:

⁴⁹ E can be thought of as the test set, which leaves the set $(C \setminus C_1) \setminus E$ to be the out-of-sample test set.

$$GGR = \frac{e_c / e}{\left(\frac{t_c}{t + e} \right)}. \quad (2)$$

Due to the automation capability of computational experimentation, the entirety of each available data set was used.

$$\text{Hence, for the case at hand, } e_c = k_c, e = k, \frac{c_1}{c_1 + e} = \frac{c_1}{c} = \frac{t}{t + e} = \frac{t}{c}.$$

Therefore,

$$GR \equiv GGR \quad (3)$$

In addition to this performance measure, statistical significance of results will be based on a significance target of 0.05⁵⁰, a power target of 0.8⁵¹, and a target confidence-interval of 95%. Each finalist network is put through a statistically sufficient number of random initializations (and subsequent training and testing) using truly-random (atmospherically generated)⁵² sequence of network-initialization seeds.

⁵⁰ If $\alpha = 0.05$, then $p < 0.05$ means the likelihood that the networks are of indifferentiable generalization quality (in statistical lingo, that their output values come from the same population) is less than 5%, which is considered a sufficiently slim chance as to consider the networks different in terms of their performance.

⁵¹ This means that the chance of being able to detect a difference (of a certain magnitude) by doing experiments is 80%. The possible magnitudes are known as small, medium, and large. The statistical literature has established minimum numbers of repetitions required for each choice of effect size, (typical) target significance, and (typical) target power. For example, in a four-way ANOVA, to set up sufficient experiments to detect a small effect with 80% power at a 0.05 target, 274 data points are needed [70].

⁵² Multiple atmospheric random-number sequences from www.random.org were used in arbitrary combinations in order to cover the required number range with no more than the required number of seeds. These are listed in Appendix O.

2.3 Artificial Neural Networks

2.3.1 Historical Background on Neural Nets

Artificial neural networks (ANNs) are interconnected structures of simple processing units known as artificial neurons. There is a wide variety of ANNs, but the most commonly used neural-net type (or, at least, the best known outside the field) is the Multi-Layer Perceptron (MLP), alternatively called feedforward neural network, *backprop* neural network, or feedforward backpropagation neural network.

Neural networks are very flexible learning agents. By flexible, it is meant they are robust in the presence of noise, and can be applied to almost any problem domain or data type. By learning agents, it is meant they can infer a mapping inherent to a given problem domain by observing examples from that problem domain.

The success rate at which they do this can be improved using a variety of techniques. Neural networks do not require the developer to program any aspect of the target function into the solution: given proper training, the network learns the target function from examples.

Training time can range from seconds to days depending on the problem domain, the amount and representation of data, and the implementation of the network. Once trained, however, neural networks are orders of magnitude faster when it comes to evaluating new examples.

Neural networks generalize well to unseen data (that is, unseen during training). Where polynomial fits can have trouble extrapolating to new data, and where linear discriminants are too simple for the mapping sought, neural networks are often the right fit⁵³. In other words, neural networks are capable of discerning the underlying structure⁵⁴ of the problem at hand. Having said that, to effect this capability into actuality, it is necessary to train neural networks with as good a coverage of the I/O space of the problem at hand as possible: no learning agent—artificial or not—should be expected to perform generalization beyond the extent to which it is provided with information about the problem domain. Hence, training neural networks and other machine-learning agents is part of the design process.

Other important aspects of the design process are the choice of features and representation of inputs and outputs. Neural networks work best with appropriately selected, compact representations that incorporate the effects of many small indicators or clues [73, p. 60]. Under such conditions, they have been shown to be reliable at

⁵³ While neural networks are said to suffer from getting stuck in local minima, this can be avoided by using good design techniques such as stochastic approximation to gradient descent and heuristic use of momentum terms (making weight changes dependent on past weight changes to effectively reduce the learning rate over time). It is, however, true that the avoidance of local minima is not guaranteed by these methods. Evolutionary algorithms attempt to avoid local minima by spreading their search throughout the evolutionary equivalent of the weight space through random mutation, but while their chances of finding the global minimum are greater, those methods also cannot guarantee the best solution for NP-complete or NP-hard problems. Both biologically inspired methods are used in practice because they can lead to solutions that are *good enough* for practical considerations.

⁵⁴ This could be anything from a mathematical function to a discrete classification to a fuzzy inference. Examples of underlying structures may include a polynomial for function fitting, chaotic dynamics for time-series prediction, classification of generalized objects for pattern detection, characteristics of events in high-energy physics, a particular type musical event for audio classification, possible states of a factory for industrial control, correlations among variables for demographics and market research, metabolic processes for biomedical signal processing, and patterns of gene expression for genomic-signal processing.

pattern recognition, detection⁵⁵, and classification; i.e., “good at Frisbee[®], bad at logic” [73, p. 59].

The implication in that slogan is that artificial neural networks provide good (and often inexplicable) solutions to problems that appear simple, but are incredibly complex, and whose human solutions are also at least partially inexplicable. Human intelligence bypasses explicitly defining the mathematics underlying problems like playing Frisbee or recognizing clave direction, and replaces it with a remarkable integration of experience, intelligence, and sensory processing. Artificial neural networks are employed in the same manner. Though particular ANNs are not general-purpose thinking devices like the biological brain, they can embody implicit knowledge about their individual problem domain, focusing successfully on the “microstructure of cognition” [75, p. 19], not the holistic mental dimension of consciousness. To put it another way, “a properly structured network can support important kinds of ‘subconscious’ reasoning that are not directly representable in the linear form of logic.” [76]

Suitable problems for neural networks are those with sufficient training sets (meaning uniform coverage of the I/O space), where long training times can be tolerated, and it is not necessary for humans to understand the learned hypothesis.

⁵⁵ Pattern recognition (pattern matching) and detection may be defined, contrastingly, as the search for frequently occurring (desirable) and infrequently occurring (anomalous) patterns, respectively [72]. This distinction is based, on one hand, on a loose definition of the term ‘pattern’ as “a configuration of data values which is of special interest,” and on the other, according to a more rigorous detection-oriented definition as a vector indicative of a local tendency divergent from the statistical baseline descriptive of a system’s behavior [72].

However, if explicit human understanding is required, decision trees can be employed in conjunction with neural networks to provide an explanation of the classification carried out by an ANN. While this is a duplication of effort, the redundancy can be leveraged by treating the NN and the decision tree as parts of an ensemble-learning system.

Also, because the reasons for a particular classification are not easily known or understood, extensive testing after training is a requirement of neural-network design.

Prestructuring, on the other hand, promises to alleviate one or more of these shortcomings of neural networks, as described below.

Research in artificial neural networks was pioneered by the idea of the artificial neuron, generally credited to psychiatrist Warren McCulloch and mathematician Walter Pitts in 1943 [77, 78]. The concept has been developed further by Rosenblatt [79], Widrow and Hoff [80], and others.

In 1949, psychologist Donald Hebb's observation that the connection between neurons (in the brain) is made stronger the more those neurons fire in concert [26] led to one of the earliest principles of ANNs: the Hebbian rule, which gave rise to the weight-update equation: $\Delta w_{ji} = \beta \times f(o_i) \times g(o_j)$, where β is a proportionality constant, and f and g are arbitrary functions of the outputs of the two artificial neurons, where o_i and o_j are the outputs of neurons i and j , respectively.

Within the following two decades, significant preliminary developments came in the form of Rosenblatt's *Perceptron* and Widrow and Hoff's *Adaline* (Adaptive Linear

Element) and *Madaline* (Multiple Adaline) paradigms, though the field has mostly moved on from these early examples and their limitations. Today, artificial neural networks are part of the technological landscape and the topic of annual IEEE conferences, with successful commercial applications to detection problems ranging from breast cancer and radar/target identification to Internet filtering and bomb sniffing.

In 1957, Rosenblatt introduced the Perceptron, which was a single artificial neuron with a specified activation function, a fixed bias input, an arbitrary number of feature inputs, and a training rule. This constituted the first fully defined neural-network *paradigm*, although a single element is a (mathematically) trivial network. The Perceptron activation function is given in (5)

$$\sum_{i=1}^n w_{ji}x_i - \theta = 0 \quad (4)$$

In the field of neural networks, a paradigm is a particular specification of three important characteristics: element type (characterized by an activation function that governs the relationship between inputs and output), connection scheme (architecture), and learning rule.

For the Perceptron, the type of artificial neuron used has a stair-step (threshold) activation function, meaning that the neuron fires only if the sum of its inputs surpasses a threshold. This is clearly a simple model of a neuron, and today serves mainly a descriptive role in teaching and learning basic neural-net concepts.

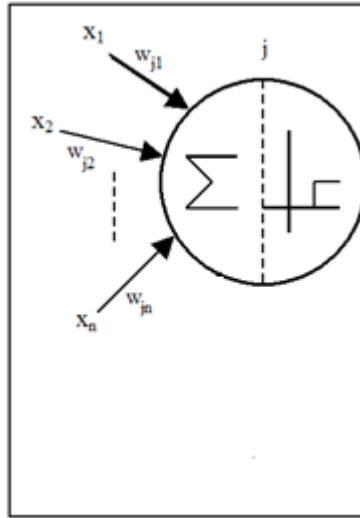


Figure 16: The Perceptron

The Perceptron Learning Rule (5) sets the change in the i -th connection weight to the product of the i -th input and the desired output. Simple exercises can show this rule to be surprisingly effective. Because the Perceptron uses a stair-step (signum, Heaviside) activation function, and is a linear discriminant, weight change that is proportional to the desired output is enough to move the network in the direction of the desired output.

$$\Delta w_i = x_i d(x) \quad (5)$$

Perceptron and Adaline inputs are connected to the element, whose output constitutes the output of the neural “net.” The learning rule for the Perceptron is known as the *Perceptron Training Rule*, followed later by the now-employed *Delta Rule*, initially developed for the Adaline of Widrow and Hoff. The Delta Rule [adapted from

81], expressed in terms of the output of the j -th element and the P -th training vector (treated as the i -th element when the network input is the only previous layer) is given by (6), where x^P is a particular training vector for pattern P , t^P is the target (desired) output for the given element-vector combination, o^P is the actual (observed) output, and η (*eta*) is the learning rate, which is the fractional constant that modifies the step size for each weight change⁵⁶.

$$\Delta w_{ji} = -\eta \frac{\partial E}{\partial w_{ji}} = \eta \sum_P (t_j^P - o_j^P) x_i^P \quad (6)$$

The Delta Rule implements ‘gradient descent’ in multi-layer (as well as single-unit, or arithmetically trivial) networks by calculating the vector derivative (gradient) of a network output’s error with respect to the current weight vector (for a single unit), or current set of weight vectors (for a network proper). This can be implemented in batch mode by considering all training examples before making any weight updates, or via stochastic approximation by updating weights after the presentation of each training example to the network.

The batch-mode error function is given by (7), where S is a training set.

$$E(w) = \frac{1}{2} \sum_{x^P \in S} (t^P - o^P)^2 \quad (7)$$

⁵⁶ In advanced implementations of neural networks, such as the NeuralWare NeuralWorks package used for the preliminary results given later in this dissertation, a “learning schedule” may be used in approximation of simulated annealing to avoid network paralysis. Such a learning schedule varies η (and other parameters) at the end of each epoch (batch size) by heuristically determined amounts, which can be altered by the designer.

The Delta Rule was originally developed for the Adaline, which, like the Perceptron, is a single-element “network,” but has a linear activation function during training. This means that the output of the element is a linear combination of its inputs, scaled based on the slope of the activation function. In operation (after learning), the Adaline is typically switched to a threshold activation function.

The Delta Rule sets the change (“delta”) in the connection weight from element i to element j via the product of the i -th element’s output and the *error* at the j -th element’s output.

2.3.2 The General Algorithm for the Delta Rule

1. Start with random weights.
2. Select an input vector from the training set.
3. If the output is not equal to the desired output, modify the connection weights as specified.
4. Go to step 2 until terminating condition.

Application of the Delta Rule moves the network in the direction of the desired output because the weight change is a fraction of the negative of the gradient (vector derivative) of the error surface. The error surface is a (possibly multi-dimensional) mapping of network-output errors to the weight vectors that give rise to them. One can appreciate that moving down the contour of an error surface in search of the smallest possible error is a good strategy as long as the surface is not multi-modal (with multiple optima). Once gradient descent begins its progress downhill, there is no reason for it to

turn back uphill, so if the valley it is in does not lead to the global minimum of the error function, the network will not move beyond the first locally minimum error that it finds.

Stochastic approximation to gradient descent (also called on-line learning) is a generally successful, but not guaranteed approach to address the local-minimum problem. While there is a single error surface per epoch in true gradient descent, the stochastic approximation seeks to avoid getting stuck in local minima by calculating the gradients on a series of different surfaces, one for each vector presented to the network (Figure 17).

When η is sufficiently small, stochastic approximation to gradient descent (updating weights after each vector, rather than after all vectors have been encountered once) closely approximates true gradient descent while reducing the likelihood of being stuck in local minima.

The Perceptron and the Adaline are mostly early developmental steps in artificial neural networks. The true strength of the ANNs, as listed at the beginning of this section, comes from the use of multiple artificial neurons in conjunction.

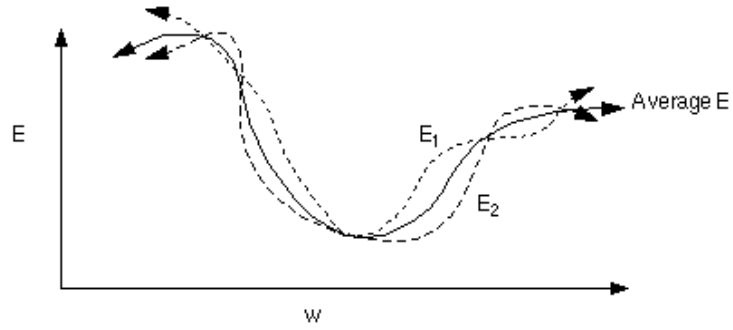


Figure 17: Changing error surfaces (shown in two dimensions) for stochastic approximation to gradient descent [82].

A single layer of multiple neurons makes for a good illustration of the derivation and operation of the Delta Rule (Figure 18):

The output of a neuron with a linear activation function is given by (8).

$$o_j^P = \sum_{i=1}^n w_{ji} x_i^P \quad (8)$$

Taking the partial derivative of both sides with respect to a connection weight, we get (9).

$$\frac{\partial o_j^P}{\partial w_{ji}} = x_i^P \quad (9)$$

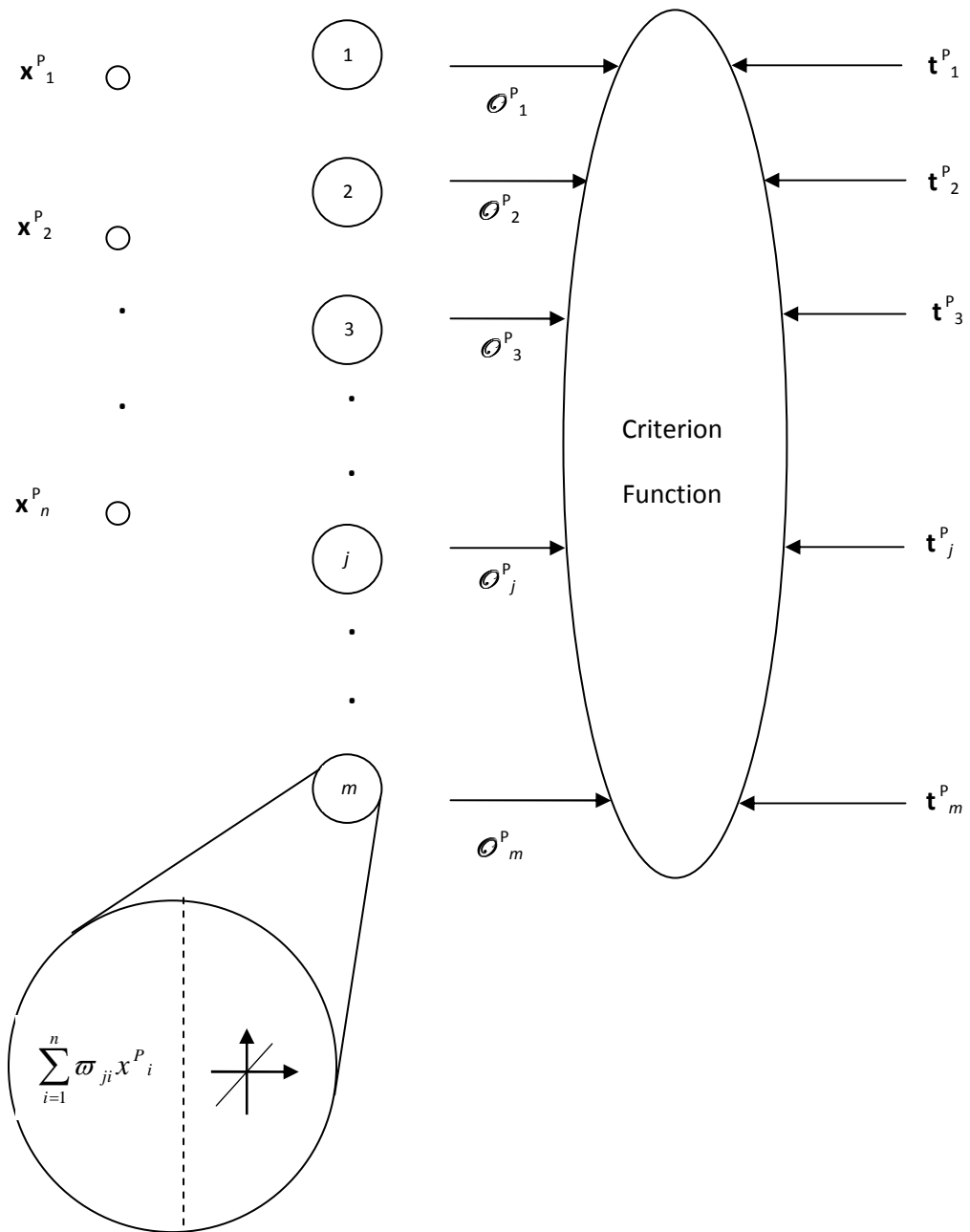


Figure 18: Single-Layer Neural Network

The criterion function, or error function (expressed for an arbitrary pattern P), is proportional to the sum of squared errors (10). Note that the summation is over the output elements.

$$CF^P = E^P = \frac{1}{2} \sum_{j=1}^m (t_j^P - o_j^P)^2 \quad (10)$$

We need to know how the error function varies with respect to the weights (11).

$$\frac{\partial E^P}{\partial w_{ji}} = \frac{\partial E^P}{\partial o_j^P} \frac{\partial o_j^P}{\partial w_{ji}} \quad (11)$$

However, by definition, the last term in (11) is equal to x_i^P . The first term on the right side of (11), the partial derivative of the error function with respect to each output, is an expression of how the error varies with the output (12).

$$\frac{\partial E^P}{\partial o_j^P} = -(t_j^P - o_j^P) \triangleq -\delta_j^P, \quad (12)$$

where $-\delta_j^P$ is the error for the P -th pattern, and pointing uphill in weight space.

Hence, we have (13).

$$\frac{\partial E^P}{\partial w_{ji}} = -\delta_j^P \times x_i^P, \quad (13)$$

and in order to go downhill, we simply take the negative of both sides (14), which is an example of the Hebbian rule.

$$-\frac{\partial E^P}{\partial w_{ji}} = \delta_j^P \times x_i^P \quad (14)$$

For the criterion function of sum-squared errors, gradient descent (the search for the lowest error in a local region of weight space) turns out to be the same as the Hebbian rule. Based on this method of gradient descent, its agreement with Hebb's discovery, and the available data and parameters $(\mathbf{x}^P, \mathbf{O}^P, \mathbf{t}^P)$, we can choose a training algorithm as in (15).

$$\Delta^P w_{ji} = \eta \delta_j^P \times x_i^P, 0 < \eta < 1. \quad (15)$$

This says to change each weight proportional to the negative of the derivative of the error with respect to the corresponding weight, as measured for the current I/O pattern. Note that the error surface for one pattern, or vector, is not the same as that for another. For a given pattern and for a specific instance of the weights, we get a point in the weight space. When weights change, we get a new point in the weight space, *not* a new weight space, as long as the same pattern is at the input. However, as other patterns P are introduced, the weight surface does change with each vector because for a new pattern, even the same set of weights defines a new error surface.

Hence, stochastic approximation to gradient descent (a.k.a. incremental gradient descent) makes use of this morphing of the error surface in weight space to attempt to avoid local minima *by switching training patterns without waiting for the training algorithm to descend all the way down a hill*, thus ***distributedly descending many gradients at once***. This random switching among weight surfaces can help avoid local minima, especially

with designer/developer vigilance⁵⁷. Stochastic descent is also faster because only one training example (or a few) is considered per descent step. However, with all these measures, stochastic gradient descent can still be subject to local minima; no guarantee of optimality is made. Increasing the number of hidden units *initially* improves the chances of avoiding local minima, but this increases complexity *and* exhibits diminishing returns beyond a point.

The Delta Rule is the same for both methods; so is the error δ . The error E is different, however, as seen in (16).

$$E = \sum_P \frac{1}{2} (t^P - o^P)^2 \quad (16)$$

For true gradient descent, P indexes all the available training patterns. For stochastic gradient descent, we have the variation shown in (17).

$$E = \frac{1}{2} (t^P - o^P)^2 \quad (17)$$

The number of hidden units is not an arbitrary parameter; it determines the expressive power of the network: Fewer hidden units are sufficient to learn smoother functions, and more units are needed as the functions sought become less smooth. Hence, playing with a hidden layer's number of units has ramifications other than the avoidance of local minima. Furthermore, multi-layer ANNs trade flexibility for certainty: While, in principle, they can represent any target function, they have no

⁵⁷ By saving, and if necessary, reverting to promising network configurations and instances along the search path.

guarantee of convergence to a global minimum-error weight vector. It is because of tradeoffs like these that the design of neural networks is an art as well as a science, and that they find use in situations where finding a practical *good* solution to a complex problem is more important than finding the *best* solution.

The above derivation for the Delta Rule applies to both the Perceptron Training Rule and the Widrow-Hoff Adaline Delta Rule because the difference between the two rules' weight-change expressions is only an additive term in δ . For the Perceptron, δ is equal to the desired output, but for the Adaline, to the desired output *minus* the actual output. Substituting one definition of δ for the other results in no change in the derivatives.

For *not-linearly-separable* problems, a single-layer network cannot perform the desired mapping; i.e., there does not exist a weight vector for the minimum error. Hence, training cannot “converge.” This was proven in 1969 by Minsky and Papert [83], who also showed that a fully connected internal layer could enable the network to represent any mapping for binary inputs. However, they made the assumption (later disproved) that there would be no convergence theorem, and thus, no extension of the virtues of the Perceptron and Adaline to the multi-layered case. This put the field of Neural Networks on hold until the mid-1980s.

The development of a credit-assignment algorithm (called *back-propagation*) for individual connection weights and the subsequent proof of universal function approximation [84, 85] (though no convergence theorem) eventually paved the way for

Multi-Layer Perceptrons (the workhorse of the Neural Networks field) and all the other paradigms of neural networks currently in use in myriad domains of science and industry⁵⁸. Derivation of the multi-layer credit-allocation rule called backpropagation is a similar but more detailed process.

2.3.3 A Brief Survey of Other Neural-Net Types

Contrary to widespread perception outside the field, neural nets are not limited to the Multi-Layer Perceptron (MLP) paradigm. MLP is only one of many supervised NN paradigms. Others include Radial Basis Functions (RBF), the Probabilistic Neural Network (PNN), the Generalized Regression Neural Network (GRNN), recursive associative neural nets such as the Hopfield and the Bidirectional Associative Memory (BAM), almost all conceivable two- and three-way hybrids of neural nets, Bayesian learning, cellular automata, evolutionary algorithms, fuzzy logic and swarm intelligence, and according to some classifications, even Support-Vector Machines (SVM), and a variety of ensemble techniques that may or may not combine NNs with other methods.

In addition, there is a family of unsupervised neural nets, the main representatives of which are the Learning Vector Quantizer (LVQ) and the Self-Organizing Map (SOM).

⁵⁸ After the resurgence of Neural Networks owing to these developments, Minsky continued to work and publish in the field. He extols the practical virtues of ANNs, but argues that they cannot achieve “higher-level reflective” thinking [86]. In the opinion of the present author, the emergence of reflective thinking, though it *may* never be brought about artificially, should not *necessarily* be considered out of reach. It may require (at least) a proper (or even, fortunate) combination of memory, recursion, hierarchical organization, and massive parallelism and complexity.

The semi-supervised family consists of different implementations of Reinforcement Learning, which are generally referred to as Adaptive Critics. (The term ‘semi-supervised learning’ is also sometimes used to mean the application of supervised learning to incompletely labeled data sets [87–89].)

Furthermore, there is a group of neural nets that combine aspects of supervised and unsupervised learning. These include the Adaptive Resonance Theory (ART) family (ART-1, ART-2, Fuzzy ARTMAP, etc.) and the Counterprop, which combines a (supervised) MLP with an (unsupervised) LVQ,

There are also refinements of the MLP, such as information maximization using overlapping receptive fields as in Linsker’s Infomax network [90], the Cascade-Correlation Learning Network (CCLN) [91], which is an algorithm for a self-growing MLP, and implementations for coarse coding like the Cerebellum-Model Articulation Controller (CMAC)⁵⁹ [92].

Neural nets can also be classified as static, dynamic, recurrent, and incremental. MLP is static; CCLN is incremental; Hopfield is recurrent. Strictly dynamic NNs are related to cellular automata, as well as other techniques of topology modification [93, 94]. Brief descriptions of a few of these paradigms follow.

⁵⁹ In addition, there are the historical NN paradigms described above that are not much in use today: Perceptron, Adaline, and Madaline.

2.3.3.1 Probabilistic Neural Network

In the family of supervised NN paradigms, but on the opposite end of the spectrum from the MLP in terms of the speeds involved in training and performance is the Probabilistic Neural Network (PNN), a neuronal implementation of what is known in statistics as kernel discriminant analysis.

For each output class or category, a probability distribution function (*pdf*) hypersurface is created via Parzen's method of adding up unit Gaussian curves centered at each feature-vector. Parzen [95] showed that with a large training set, the true *pdf* of the underlying function is approached. Specht showed that classification accuracy is relatively insensitive to the choice of variance, as long as it is small or intermediate (not much greater than 2) [96].

To reduce computational load, a Taylor-series polynomial expansion can be used to approximate the hyper-*pdf*s for each class. This approximation of an approximation actually serves to reduce occurrences of overfitting. If it had not been for the smoothing effect of Parzen windowing, the Taylor polynomial would fit a precise curve represented by the training set to too great an accuracy. [96]

Furthermore, no matter the dimensionality of the actual feature vector, the polynomial expansion can use as many or as few parameters as desired, independent of the number of features. There is a trade-off between over-smoothing and overfitting, which comes with corresponding computational and memory costs. In other words, using too few parameters in the Taylor approximation may lead to occasional large

classification errors. These can be avoided with higher-order approximations, which take up memory and computation time. However, the probability distribution function is not known in advance and the appropriate choice of order may require heuristics, experimentation, or insight.

2.3.3.1.1 Architecture of PNN:

1. The Distribution Layer cardinality (number of input nodes) is the number of *features*.
2. The Pattern Layer (hidden layer) elements are as many as the size of the training set. There is a neuron for each training vector, and the *instar* for that neuron (its vector of connection weights) is set equal to the values of that training vector.
3. The Summation Layer (output layer) has as many units as there are classes. The neurons in the Pattern Layer are grouped into the Summation Layer neuron for the corresponding class.
4. The Decision Layer makes a comparison among incoming Summation Layer outputs.

2.3.3.1.2 Advantages of PNN:

1. No training time is required beyond the time needed to read in the data and run the network once.
2. Training data can be added or removed without additional training [96, p. 36].

3. With enough data, PNN approximates optimal (not naïve) Bayesian classification [96, pp. 35 and 53], which is optimal on average in that no other method can have statistically higher prediction accuracy [96, p. 37].
4. With sparse data, PNN performance only degrades to a nearest-neighbor classifier [96, pp. 43–5], which is a simple and popular classification method.
5. PNN provides an indication of the amount of evidence for its result (confidence) [96, p. 36].
6. PNN can learn arbitrary, complicated relationships with nonlinear decision boundaries [96, pp. 35–6].
7. It is largely noise-immune.
8. It features no network paralysis.
9. It does not get stuck in local optima.
10. It is fully parallel.

2.3.3.1.3 Disadvantages of PNN:

1. It requires as many units (artificial neurons) as the number of training patterns (in the internal, or pattern, layer) *plus* the number of inputs (at the input, or distribution, layer) *plus* the number of outputs (at the output, or summation layer) [96]; hence many units for large training sets or long feature vectors.
2. Vectors that are scalar multiples of one another are indistinguishable.
3. Although PNNs train in almost no time, they can take disagreeably long to perform each classification. This is the exact opposite of the MLP, which takes more time for training than for classification (generalization).

4. As an implementation of kernel discriminant analysis, the PNN interpolates successfully, but does not extrapolate—i.e., generalize—well outside the section of space covered by the training set.

2.3.3.2 Unsupervised Networks

Unsupervised networks operate on the basis of competitive learning. The typical examples of this are Kohonen's Learning Vector Quantizer (LVQ) and Self-Organizing Map (SOM), whose weights are initialized randomly and then compete for the closest match to an input pattern. In the simplest implementation of competitive learning, the weight vector that is closest to the input vector is the sole “winner” and is moved closer to the input vector. This continues for each input pattern, eventually placing weight vectors at the centroids of clusters of input vectors. Note that for this to work, the weight space and vector space have to be superimposable, hence of the same cardinality. A number of improvements, such as a “conscience” factor to keep a single vector from always winning, have been made to the original LVQ.

2.3.3.3 Associative Memory

Memory in computer applications and design has mostly been based on the notion of ‘place’. In modern general-purpose computers as well as dedicated embedded controllers, each item in memory is stored in a particular memory location and accessed via its address. Both hardware and software follow this model for the vast majority of modern computing. [97, p.3]

Similar ideas have been put forth in the psychological study of human memory, but over the last several decades, models of distributed, associative and content-addressable memory have won over the centrally managed location-based model of memory for understanding and mimicking the biological brain.

Consequently, such associative memories have been developed in several forms in the field of Neural Networks. These include Hopfield Networks, Bidirectional Associative Memory (BAM), Linear Associative Memory and Fuzzy Associative Memory.

The Bidirectional Associative Memory (BAM) can employ either binary or bipolar-valued data. The network is *hetero-associative*, meaning different patterns x and y (not necessarily of the same size) are associated with one another through the bidirectional network connections (See Figure 19).

A simple example of the operation of associative memories is given below.

$$n = 3, m = 4$$

$$\text{Pair 1: } \mathcal{A}_1 = [1 \ -1 \ 1] \text{ and } B_1 = [1 \ -1 \ 1 \ -1]$$

$$\text{Pair 2: } \mathcal{A}_2 = [-1 \ 1 \ 1] \text{ and } B_2 = [-1 \ 1 \ -1 \ 1]$$

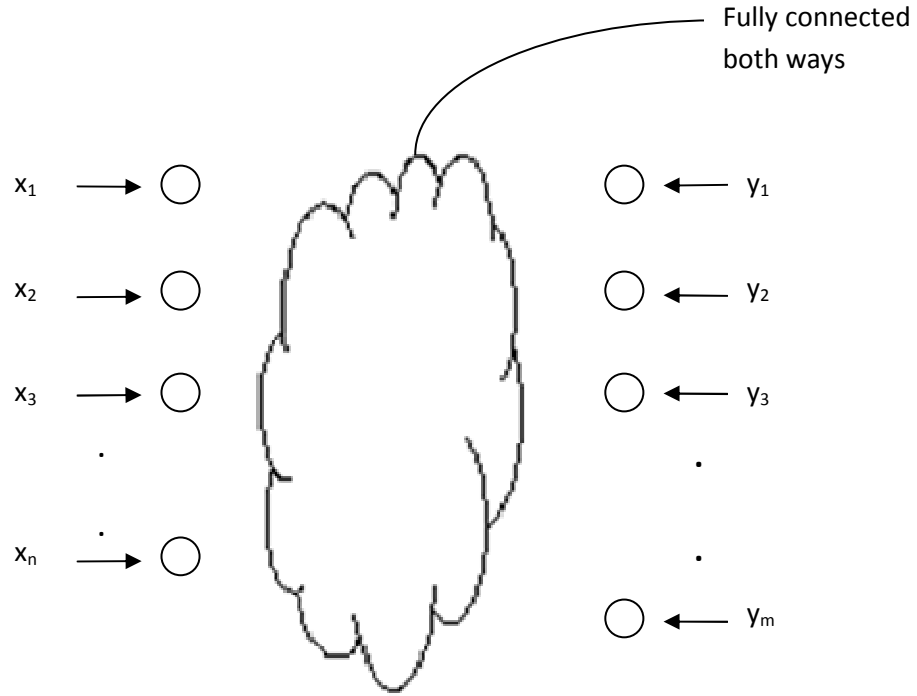


Figure 19: Basic Bidirectional Associative Memory (BAM) setup

A weight matrix is created out of the first pair:

$$A_1^T B_1 = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} [1 \quad -1 \quad 1 \quad -1] = \begin{bmatrix} 1 & -1 & 1 & -1 \\ -1 & 1 & -1 & 1 \\ 1 & -1 & 1 & -1 \end{bmatrix} = W_1 \quad (18)$$

and another from the second pair:

$$A_2^T B_2 = \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix} [-1 \quad 1 \quad -1 \quad 1] = \begin{bmatrix} 1 & -1 & 1 & -1 \\ -1 & 1 & -1 & 1 \\ -1 & 1 & -1 & 1 \end{bmatrix} = W_2 \quad (19)$$

The current conception of associative memory is that data are not stored in any particular discrete location, but *everywhere* [98, p. 3], just as spread-spectrum technology combines many streams of data, or a high-Q filter extracts individual components from a complex signal. Hence, in the current simple example, the two matrices are added together:

$$W = W_1 + W_2 = \begin{bmatrix} 2 & -2 & 2 & -2 \\ -2 & 2 & -2 & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad (20)$$

Associated vectors can then be recalled through further matrix multiplication and a thresholding rule. When one of a pair is presented, that vector is multiplied by the memory contents (the sum matrix) and the result is thresholded as follows:

$$b_j(t+1) = \begin{cases} 1 & \text{if } y_j > 0 \\ \text{same} & \text{if } y_j = 0 \\ -1 & \text{if } y_j < 0 \end{cases}, \text{ where } y_j = \sum_{i=1}^n a_i(t) w_{ji}, \text{ as in the}$$

following recall example.

To recall the associated pattern for the vector $[1 \ -1 \ 1]$, first calculate $A_1 W$:

$$A_1 W = [1 \ -1 \ 1] \begin{bmatrix} 2 & -2 & 2 & -2 \\ -2 & 2 & -2 & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix} = [4 \ -4 \ 4 \ -4] \quad (21)$$

which after thresholding according to the $b_j(t+1)$ rule: $B_1 = [1 \ -1 \ 1 \ -1]$. The associated vector is recalled.

Of course, the main interest in associative networks is when incomplete or noisy data are presented. In that case, the thresholding rule, or the error surface associated with the related error function, enables pattern completion.

Let A'_1 be $[0.7 \ -1.1 \ 0.8]$.

$$A'_1 W = [0.7 \ -1.1 \ 0.8] \begin{bmatrix} 2 & -2 & 2 & -2 \\ -2 & 2 & -2 & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix} = [3.6 \ -3.6 \ 3.6 \ -3.6],$$

which after thresholding gives $[1 \ -1 \ 1 \ -1]$, which is the associated vector we wish to retrieve.

In summary, the example above illustrates how associative memories store multiple patterns in distributed fashion by storing one matrix sum of outer products of pairs of data vectors to be associated and recalled.

2.3.3.4 Counterpropagation

The *counterprop* network [99] was originally designed with both forward and backward propagation (much like the biological nervous system), but has since come to mean a network that combines an LVQ layer, which serves as preprocessing (feature extraction), with an MLP layer that is turned on after sufficient clustering by the LVQ layer. This paradigm may be used when training-set classification is uncertain or unknown, and preprocessing of the training data required to reduce data complexity cannot be done solely on the basis of domain knowledge.

2.3.3.5 Adaptive Resonance Theory

Similar to Counterprop in its applicability, Adaptive Resonance Theory (ART) [100, 101] provides a variety of methods for classifying binary, non-binary, crisp, and Fuzzy data when the number and membership for categories needs to be flexible. ART implementations can create new output categories or classify inputs to existing ones, depending on user-defined parameters such as “vigilance,” which control the level of strictness with which a pattern must belong to existing categories, or the ease with which it can be made into a category of its own. These networks implement the plasticity of biological neural nets; the designer does not have to possess knowledge of all possible categories in advance.

This wraps up the summary of some of the leading neural-network paradigms.

2.4 Reconstructability Analysis

“Reconstructability analysis (RA) decomposes ... data in the form of either set-theoretic relations or multivariate probability distributions, into ... relations or distributions involving subsets of variables.” [102]

The material in this section requires knowledge of Information Theory. The inexperienced reader is referred to Appendix B for a gentle and intuitive treatment of Information Theory.

In the proposed research, vectors consisting of 16 binary inputs and one binary or multi-valued output (depending on encoding and modularization) are used to characterize the attack-point representation of rhythms abstracted from traditional

Afro-Brazilian music⁶⁰, along with the membership that each such rhythm pattern holds in the author’s particular interpretation (and set extension) of the *carioca* implementation of a West African rhythmic device involving a call-and-response characteristic, described by the African musicologist Anku [103, pp. 172–175] as “the principle of mobility and finality,” also known as “clave sense” or “clave direction.”

From the research point of view, inquiry about this rhythmic aspect has two advantages:

It is governed by a complex, yet rarely studied perceptual and cognitive phenomenon. This phenomenon and its proposed representations allow for the formulation of a research question likely to illuminate a long-standing issue in the field of Artificial Neural Networks.

The introductory quote to this section reveals the connection between the set-theoretic nature of classifying rhythmic patterns into musical categories, and the applicability of probabilistic modeling via Reconstructability Analysis [104] to such problems. Although they are musically abstract, binary-valued attack-point rhythms are finitely enumerable and thus constitute a set-theoretic classification problem⁶¹. The

⁶⁰ The rhythm patterns can actually come from any source (Brazilian music, other music, one’s imagination, or a random-number generator). The underlying idea is that *any* rhythmic pattern is possible (though not equally probable) during improvisation. One would like to know the degree of fit of any such pattern to the tradition under consideration.

⁶¹ From the musical point of view, such a classification is not absolute (is open to debate, improvement, and interpretation) because a) the clave problem is not traditionally defined for all enumerable rhythms, so a set extension of the concept had to be deduced by the author, based on known principles of clave direction in samba *carioca*, b) the clave tradition is interpreted and implemented differently in different cultural traditions, and even within a given culture, and c) stripping away pitch, tempo, expressive timing and other musical-contextual information in creating attack-point rhythm makes each abstract vector open to multiple interpretations, and a narrow set of

actual model-search process, however, breaks the space of binary variables into proper subsets, resulting in incompletely specified functions which give rise to an information-theoretic search. This is because each incompletely specified function, say the variable chain $BCDFGJKMQR$ out of the complete set of $\{A, B, C, D, E, F, G, H, J, K, M, N, P, Q, R, S\}$, can have an occurrence count greater than one. This means that any model search on a subset of the full variable set is an information-theoretic (probabilistic) search.

In addition, exploratory modeling, by its nature, is statistical in that whenever many models are present, some will fit any data by chance. This is a problem of statistical significance: The investigator must estimate the likelihood of any apparently successful model being a truly good model or a chance model. RA includes metrics to this effect.

2.4.1 RA and *Occam3*

Reconstructability Analysis (RA) is a set of modeling and model-evaluation methods for analyzing nonlinear relationships between multiple discrete nominal variables for the purpose of evaluating trade-offs between reduced complexity and reduced information in modeling such data, and hence, the underlying system (the *whole*, in Systems Science terminology). It is primarily divided into two categories: information-theoretic RA (IRA) and set-theoretic RA (SRA), even though each type contains aspects

musical and music-teaching settings had to be selected by the researcher to make classification at all possible.

of the other. For instance, IRA is concerned with set-theoretic relations but examines them in a probabilistic rather than relational manner.

RA is applicable to static as well dynamic, deterministic as well as stochastic relations [105, p.1]. The capacity of RA to accommodate dynamic relations; i.e., variables that represent the state of a system at sequential time samples, makes it suitable for the abstract musical onset data proposed for use at the first stage of this research. Further theoretical justification for this representation can be found in the section on background and domain knowledge below.

In modeling multivariate data, the primary trade-off is between complexity and accuracy (fidelity). In most cases (discounting inconsistent and highly noisy data), the full set of available data represent the underlying function more accurately than a proper subset of such. However, working with the full set of available data may be infeasible from the point of view of human-hours, computational cost, practical realizability (as in the case of NP-hard or NP-complete problems), or some other practical concern. Also, in some cases, the complete set of data simply may be unavailable, or even uncollectable. Modeling aims to reduce data complexity, either without loss of accuracy or with an acceptable degree of such loss to make the use of complex data possible. IRA, as implemented in the application used in this study (*Occam3*, the Organizational Complexity Computation and Modeling program, <http://dmm.sysc.pdx.edu/occam/weboccam.cgi>), searches the space of either the full set of relations, or a variety of proper subsets, based on heuristics and user input that

guide the search in finding and rating models of the data by various information-theoretic metrics.

The key to “reconstructability” is model evaluation. A model is a set of relationships among variables. Different models explain the data at hand to various degrees of fidelity, and have various degrees of complexity. Clearly, the “saturated model” (defined as a direct copy of all of the relationships found in the complete data set—not a “model” at all, but a copy of the data, which is a mathematically trivial model) is the one with the greatest explanatory power for the given data set. However, it is here that we see the difference between explanatory and predictive power of a model. Overfitting, or overtraining, is a phenomenon in Machine Learning wherein a model (also known as a hypothesis) agrees with the data so well that it captures not only the general system characteristics in the data but also the particular eccentricities of that selection of data such that it cannot generalize to other instances corresponding to the same general systemic characteristics but a different chance arrangement of detailed features. (In Statistics, this is described as a model having low bias and high variance [39, p. 38].)

Whenever less than the complete data set is used to model a system, the “saturated model” is such a chance selection of features and attributes (even if the selection was not made purely by chance).

On the opposite end of the spectrum, the “independence model” (for bottom-up regular RA) or a uniform distribution (for bottom-up state-based RA) constitute the

simplest possible model (again, not a model at all, but an application of scientific honesty based on Laplace’s principle of insufficient reason [18, pp. 4–14]: “Out of all probability distributions satisfying given moment constraints ... choose the distribution that is closest to the given *a priori* probability distribution \mathbf{q} , and in case the latter is not specified, choose the distribution that is closest to the uniform distribution.” [18, p. 14]

Besides providing us with the “bottom model” or “reference model” the principle of entropy maximization (or Laplace’s principle of insufficient reason) also guides the model-selection process in RA (and in all other information-theoretic modeling techniques) by establishing that the best (most scientifically honest) model out of a selection of equally predictive/equally explanatory models is always the closest to that suggested by *a priori* information (not applicable in the present research other than being worked into the data itself), or, in its absence, the uniform distribution.

Reconstructability Analysis, then, navigates the (possibly multi-dimensional) landscape between the two extremes of trivial models, trading off complexity and fidelity. In searching up from an independence or uniform reference model, RA asks whether the “increase in complexity [is] justified given the gain in fidelity (i.e., reduction in error)” [106]. Similarly, in searching top-bottom, with the data (saturated model) as reference, RA asks whether “the loss of fidelity (i.e., increase in error) [is] acceptable given the decrease in complexity” [Ibid.].

In both cases, complexity is measured by “degrees of freedom.” Abbreviated *df*, the degrees of freedom of a system can be exemplified by a simple algebraic equation

such as $x + y + z = 12$. This equation, also known as a constraint, narrows down the possibilities for the values the variables x , y , and z can take. In the absence of other constraints, one can pick any one variable with utmost freedom. For instance, let's arbitrarily pick z to be 37. We can also arbitrarily pick the value of one other variable, say x . However, once we pick a value for x as well, we are locked in to one and only one value for y . We do not have the freedom to choose y —it is determined by our previous choices. If $x = 5$ and $z = 37$, then y has to be -30 . This system has two degrees of freedom, one less than the number of variables.

Clearly, a more complex system, such as $t + u + v + w + x + y + z = -1$, has more degrees of freedom, and a less complex system has fewer. As in these simple algebraic expressions, the data set presented to an RA search constitutes a set of constraints on the values variables can take in each of their relational arrangements. With more complex and interacting constraints, as in the data for this research, calculations of degrees of freedom are more complicated, but follow from the same basis.

Degrees of freedom in RA results are given as Δdf , the change in degrees of freedom with respect to the last winning model in the search. This is known as an “absolute” search criterion. Some RA criteria can be calculated both absolutely and “incrementally.” While the incremental calculation is a simple subtraction with degrees of freedom, it is considerably more complex and more useful for statistical criteria like the α significance level.

Some simple (two- and three-variable) examples of the operations of projection (analysis) and composition (re-synthesis, or reconstruction) in RA is shown in the following series of figures (based on [105, 107]).

		z_1		z_2	
		y_1	y_2	y_1	y_2
x_1		a	b	c	d
		e	f	g	h

		z_1		z_2	
		y_1	y_2	y_1	y_2
x_1		$p(x_1, y_1, z_1)$	$p(x_1, y_2, z_1)$	$p(x_1, y_1, z_2)$	$p(x_1, y_2, z_2)$
		$p(x_2, y_1, z_1)$	$p(x_2, y_2, z_1)$	$p(x_2, y_1, z_2)$	$p(x_2, y_2, z_2)$

Figure 20: A 2 x 2 x 2 contingency table shown in shorthand notation (above) and explicit probability notation (below).

The Entropy for any combination of variables is given by (33).

$$H(x, y) = - \sum_i \sum_j p_{ij} \log_2 p_{ij} , \quad (33)$$

which, for the data in Figure 21, is equal to the expression in (34):

$$-0.1\log_2 0.1 - 0.2\log_2 0.2 - 0.3\log_2 0.3 - 0.4\log_2 0.4 \quad (34)$$

If we were to refer to the values whose logarithms are taken in (34) as a, b, c, d , just as in Figure 20 (top row of top half), then the shorthand notation for $H(x, y)$ is $\Gamma(a, b, c, d)$.

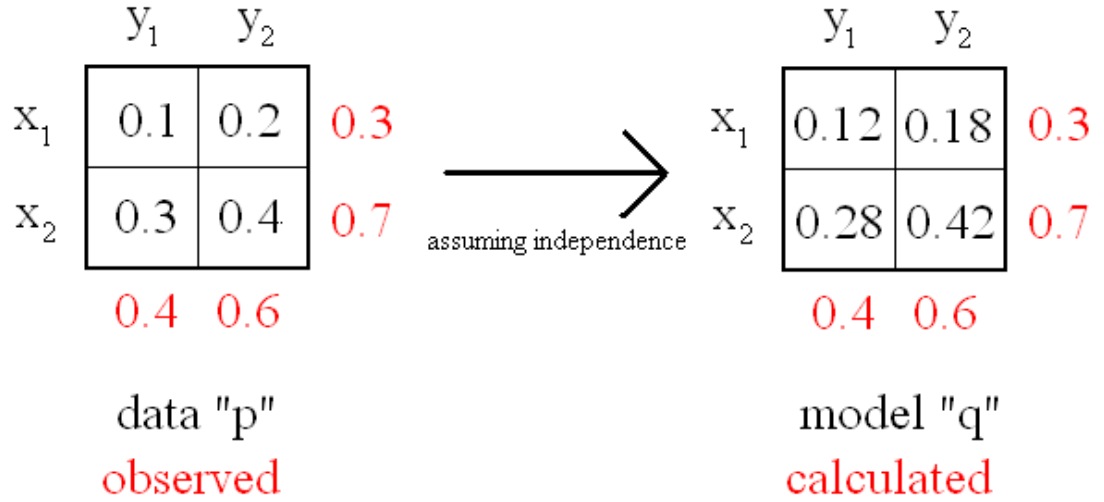


Figure 21: Example of calculated two-variable model values based on the marginals of the observed distribution (data).

In Figure 21, each cell of the calculated distribution⁶² is the result of the product (due to assuming probabilistic independence) of the marginals of the observed data—the row and column totals. The margins are preserved from “p” to “q” and give rise to the calculated distribution in the latter. Differences such as that between 0.1 under “p” and 0.12 under “q” is the Transmission in (43). The value of Transmission is a measure of

⁶² Strictly speaking, neither the observed data nor the calculated values form a distribution in the statistical sense. The former is a sample population and the latter is an estimate. However, In RA, the term distribution is used for these data, and not only for the platonic generating process.

the extent to which fidelity was sacrificed in order to gain generality. The arithmetic details follow.

The marginal for variable y is $\Gamma(a + b, c + d) = \Gamma(0.3, 0.7)$. Hence, if one were to compare the individual values of the “p” and marginal values of the “q” distributions of Figure 21, one would have (35):

$$\Gamma(0.3, 0.7) + \Gamma(0.4, 0.6) - \Gamma(0.1, 0.2, 0.3, 0.4) = H(X:Y) - H(XY) = T \quad (35)$$

where the last term is Transmission, and the second-to-last term is “the data,” equal to $\Gamma(0.1, 0.2, 0.3, 0.4)$. (Note that RA-structure notation uses capital letters whereas probability notation uses lowercase, but they follow the same definition of Entropy.)

From this modest start, we can express the Uncertainty of (Entropy in) y given knowledge of x in two ways. First, (36):

$$H(y|x) = H(x, y) - H(x) = \Gamma(0.1, 0.2, 0.3, 0.4) - \Gamma(0.4, 0.6) = 0.8755 \quad (36)$$

where the first expression is literally “the uncertainty in x removed from the uncertainty in the two-variable system.”

The second way is to take each value (state) of x separately, and calculate the weighted sum of conditional probabilities (37):

$$H(y|x) = \sum_i p(x_i) H(y|x_i) = 0.4 \Gamma\left(\frac{0.1}{0.4}, \frac{0.3}{0.4}\right) + 0.6 \Gamma\left(\frac{0.2}{0.6}, \frac{0.4}{0.6}\right) = 0.8755 \quad (37)$$

For the three-variable case of Figure 20, the various Entropy values are calculated as follows (38–42).

$$H(x, y, z) = \Gamma(a, b, c, d, e, f, g, h) \quad (38)$$

$$H(x, y) = \Gamma(a + c, b + d, e + g, f + h) \quad (39)$$

$$H(x, z) = \Gamma(a + b, c + d, e + f, g + h) \quad (40)$$

$$H(y, z) = \Gamma(a + e, c + g, b + f, d + h) \quad (41)$$

$$H(x) = \Gamma(a + b + c + d, e + f + g + h) \quad (42)$$

and so on. Referring again to Figure 21, if Transmission is greater than zero, there is an association between the observed and calculated distributions; if $T = 0$, there is no association. Now consider (43) which states that Transmission of the model is equal to the Uncertainty of the model ($H(X:Y)=H(x) + H(y)$, the margins) minus the Uncertainty in the data ($H(XY)=H(x,y)$, the joint distribution), which is necessarily smaller. This quantity is the degree of divergence of the data from being uniform (maximum entropy).

$$T(X:Y) = T(q) - T(p) = H(X:Y) - H(XY) = \Gamma(0.12,0.18,0.28,0.42) - \Gamma(0.1,0.2,0.3,0.4) = 0.0058 \quad (43)$$

RA promises to provide information for guiding neural-net architectures in a number of ways. It can find relationships among variables, leading to a divide-and-conquer type of prestructuring where a number of smaller neural nets learn simpler mappings. It can also lead to the elimination of variables whose contributions to classification are judged low enough to be disregarded. More importantly, RA can

discover interaction effects among variables, guiding what inputs should be connected to what hidden-layer elements.

Interaction effects are when the predicting power of variables changes when those variables are considered together as well as separately. A purely hypothetical but easy-to-conceive example may be lethal bicycle accidents in big cities. Assume that extensive data is available that identifies two factors (independent variables): type of helmet and speed of colliding car. Each variable alone has a certain predictive (explanatory) power with respect to the fatality rates in bicycle–automobile collisions; i.e., the helmet variable alone can predict death rates for collisions within a given range of speeds at a certain level of accuracy. Similarly, the speed variable also has its own level of explanatory power. However, if there are interaction effects, these variables considered separately will never explain the observed fatality rates (or predict future fatality rates) as closely as when interactions are also taken into account.

For instance, let the levels of the helmet variable in the data be *no helmet*, *cheap helmet*, and *good helmet*. Similarly, let the automobile speeds be *low speed* and *high speed*.

The null hypothesis would be that neither variable has any effect on the outcome. If this were true, we would see a table of factors (inputs, independent variables) and outcomes (outputs, dependent variables) that looks like Table 2 where neither helmet type nor auto speed has any effect on fatality. This is the null hypothesis. (These data are not factual, but contrived for demonstration purposes only.)

Table 2: Data Breakdown for Null-Hypothesis Example

Fatalities under the Null Hypothesis	Variable 1 →	HELMET TYPE			Marginals ↓
Variable 2 ↓	States of Variables → & ↓	None	Cheap	Good	
AUTO SPEED	Low	15	15	15	45
	High	15	15	15	45
Marginals →		30	30	30	90

For the case where helmet type correlates with fatality, but auto speed has no effect, the data might look like Table 3.

Table 3: Single determining factor.

Speed- Invariant Fatalities	Variable 1	HELMET TYPE			Marginals
Variable 2	Variables' States	None	Cheap	Good	
AUTO SPEED	Low	24	15	6	45
	High	24	15	6	45
Marginals		48	30	12	90

Similarly, if auto speed were the sole determining (or at least, correlating) factor, the data might appear as in Table 4.

Table 4: Single determining factor; reversed correlation.

Helmet-Invariant Fatalities	Variable 1	HELMET TYPE			Marginals
Variable 2	Variables' States	None	Cheap	Good	
AUTO SPEED	Low	1	1	1	3
	High	29	29	29	87
Marginals		30	30	30	90

Finally, in Table 5, the data, as reflected by the marginals, indicate that both factors relate to the outcome, and interact in their effect. For any given row or column, the effect of different helmet (for row) and auto speed (for column) are monotonic, but when traversing the table diagonally, we now see the interaction effect that no helmet with a low-speed collision results in fewer fatalities than cheap helmets in high-speed cases.

Table 5: Output Dependence on Both Factors

Interaction Effects	Variable 1	HELMET TYPE			Marginals
Variable 2	Variables' States	None	Cheap	Good	
AUTO SPEED	Low	19	2	2	23
	High	38	27	2	67
Marginals		57	29	4	90

Returning to clave, interactions effects are very common in the *partido-alto* interpretation of clave direction used as the data source for the present research. To demonstrate this, let's start with a simple comparison and progressively increase the number of expressed onsets.

Under the firm-teacher model of samba *carioca*, the attack-point rhythm 0100|0000|0000|0000 is in category 1 (the 3-2 direction), whereas the vector 0000|0000|0000|0000 is neutral (category 3). Thus changing the value of the RA variable 'B' has altered the category of the output.

However, keeping the variable 'B' high does not guarantee membership in category 1. Setting the variable 'K' high will take us back to category 3: 0100|0000|0100|0000, and setting the variable 'H' high (0100|0001|0100|0000) will take us all the way to category 2 (2-3 direction). This is an interaction between 'H' and 'K', where neither is able to override the high value of variable 'B' to flip the clave direction of the pattern alone, but together, they do.

A finer effect of interactions can be seen in the degree of membership a pattern has in a given class direction. The variable ‘B’ is a powerful variable in the present model in that it has high predictive power. However, it interacts with the variable ‘C’ such that a 1 on ‘B’ has a much greater impact on the classification of pattern if ‘C’ is low (0100 | ...) than if ‘C’ is high (0110 | ...). This is similar to the powerful impact of the automobile speed, which is highly predictive, but whose power can be diminished or augmented by the state of the helmet variable.

2.4.2 Historical Reference

Reconstructability Analysis (RA) was developed over the past five decades, starting with the work of Ashby [108, 109], and “developed by Broekstra, Cavallo, Cellier, Conant, Jones, Klir, Krippendorff, and others” [110, 111].

2.4.3 RA Formalism

The basics of RA and associated theory are set out in detail by Zwick [44, 102]. The lattice of structures given in these papers is the complete set of possible models for a four-variable system. It should be noted that the boxes represent relations, and the lines represent variables (Figure 22).

There are 20 structure types and 114 specific structures for four variables [102]. For a 17-variable system, as discussed here, the number of models in the entire lattice of structures is prohibitive from the point of view of computational cost, though the directionality (input/output nature) of the system constrains the number of specific models to one or more orders of magnitude less than that for neutral systems (no

input/output distinction). Nonetheless, the lattice search suffers the combinatorial problem of the curse of dimensionality [25], and a combination of randomization and heuristics are proposed here to perform as thorough a search of the lattice of structures as necessary. (See Section 4.6.5 for details.)

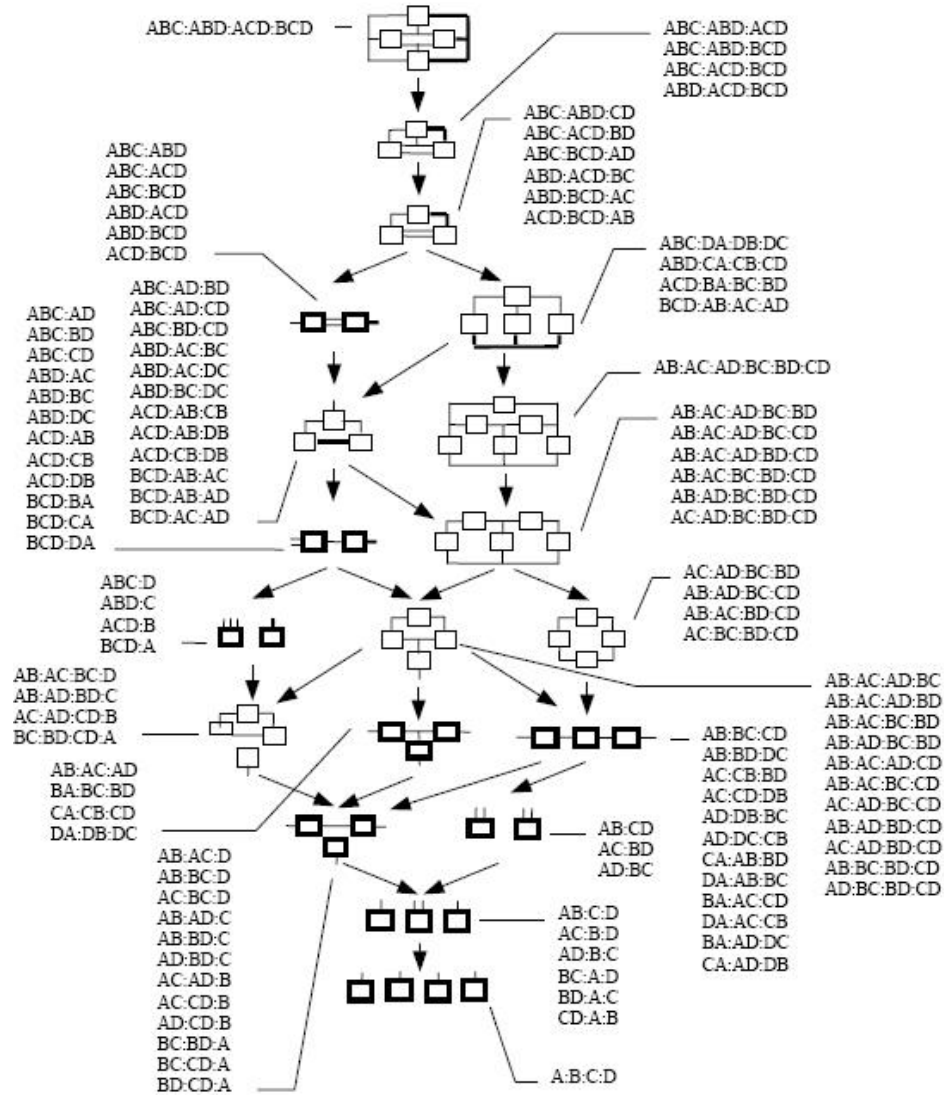


Figure 22: The Lattice of Structures for (only) four variables, taken from [47].

2.4.4 Aspects of Statistics Relevant to RA

RA is a method of quantifying relationships among variables using an extension of χ^2 (chi-squared) testing for independence.

Information is constraint captured in the model (or percent-uncertainty reduced). Constraint *lost* in a model is the error in that model, which is the distance from the data.

The null hypothesis is that “the model is as good as the reference.” Rejecting this is to claim that the model is too far from the reference. Note that the reference can vary, depending on the problem or its interpretation. It can be the bottom (independence) or the top (data). If one is wrong in rejecting that the model is as good as the reference, one has made a type-I error.

Alpha is the probability of a type-I error. Alpha is, then, the probability of being wrong in rejecting that the model is close to the reference.

In other words, alpha is the probability of being wrong when saying the model is not close to the reference. To put it another way, alpha is the probability that the model *is* close to the reference.

Given TOP: ABC (data, “saturated”) and BOTTOM: A:B:C:Z (independence), Default reference *and* starting model for directed systems (systems with causal relationships of inputs and outputs) is the BOTTOM.

Given the reference is the bottom (independence), alpha is the probability of being *wrong* when one *rejects* that the model is the same as independence.

Equivalently, alpha is the probability of being *wrong* in saying that the model is far from independence (close to the data).

Alpha is, then, the probability of being RIGHT when one says that the model is the same as the data (which is desirable).

In a χ^2 test for independence for two variables, the null hypothesis is that the variables are independent. The test results in a P value, which indicates whether the null hypothesis (independence) may be rejected or not. The smaller the P value, the stronger the evidence that the null hypothesis may be rejected, and the research question (alternative hypothesis) may be supported.

A P value of 0.01 implies that only “1 sample out of 100 would provide [convincing evidence] for the alternative hypothesis when in fact it were not true” [112, p. 492], or that the probability of observing (by pure chance) data as distant from the null hypothesis as that observed is 1%.

The upper limit for the probability of a type-I error (rejecting the null hypothesis when it is true) is set as the *significance level*, denoted by α . “We can reject the null hypothesis at a specified level of significance α only if the P -value from the sample is less than or equal to α ” [Ibid., p. 493] because this says that the probability of observing a sample such as the one we have observed by pure chance is less than the *tolerable* probability of using such a sample to reject the null hypothesis when we should

not. In other words, the significance level α is the type-I-error probability which we can tolerate. If this is the case, we say the sample is *statistically significant at the α -level*.

To express the same thing in terms of confidence intervals, we can reject the null hypothesis at the α significance level *if and only if* a $(1-\alpha)$ confidence interval does not include the null-hypothesis value [Ibid.]

Statistical significance—quite different from everyday significance—has to do with the likelihood that observed data reflect the realities of the underlying system. There is always the possibility that when a population is sampled, the observed distribution is due to pure chance (innocent sampling error beyond the control of proper design practices), and not a reflection of the underlying facts, structures, or trends. Measures of statistical significance provide a sense of how likely this is to be the case. Such measures are sometimes more meaningful if a target significance level, based on convention and the importance of rigor for the problem at hand, have been determined prior to data processing. In that case, there are two interpretations of α [27]:

If the experimenter has no *a priori* expectation, bias, belief, or hypothesis about the plausibility of a particular discrepancy, then “one begins to be slightly suspicious of a discrepancy at the 0.20 level, somewhat convinced of its reality at the 0.05 level, and fairly confident of it at the 0.01 level. [27]

If the experimenter has any *a priori* expectation, bias, belief, or hypothesis about the underlying system (which is most often the case), this must affect her/his attitude and choice of target significance level. “If the alternative hypothesis [were] plausible a

priori, the experimenter would feel much more confident of a result significant at the 0.05 level than if it seemed to contradict all previous experience.” [27]

On the other hand, the idea of statistical significance and the use of certain conventional significance levels can, and frequently is, over-used to the extent that they can be less informative than simply stating the probability of observing a discrepancy as large as that observed (or larger) by pure chance. This is because common practice has made certain levels of statistical significance conventional in certain disciplines. Among these are 0.01, 0.05, and 0.10. “The statement that a particular deviation is ‘not significant at the 0.05 level’ is sometimes found to mean, on closer examination, that the actual probability is 0.06” but the *impression* that it must, then, be 0.10 was made by unjustified attachment to conventional levels of significance [27].

In either case, it is imperative to note that statistical significance does not necessarily infer any degree of physical, social, psychological, financial, medical, or other importance. It is better understood to mean *plausibility* [27] or *rareness* [113].

In terms of the models that are considered by the *Occam3* software when carrying out Reconstructability Analysis (RA), the null hypothesis (in the cumulative sense of significance) is the starting model. Depending on the type of search performed, the starting model falls into one of four categories: the independence model (bottom-up), the saturated model (top-down from the full data set), the empty model, and the user-selected model.

RA differentiates between directed and neutral systems. Directed systems have inputs and outputs; neutral systems have no sense of input and output variables, though relationships may still exist among variables. ***The only system type considered here is directed.***

For directed systems, the default reference model in *Occam3* is the independence model (“bottom”), in which case a search of other possible models is performed with the objective of determining whether a more complex model is justifiable in terms of trade-offs involving the information gained or lost, percent accuracy, and the statistical significance of rejecting the independence model.

Once an acceptable model has been identified through the *Occam3* search feature, a more detailed search (called “fit”) may be performed by using the selected model as the reference model. In this case, the α values reported by *Occam3* are with respect to so-called *incremental* changes from the current reference model, as opposed to cumulative changes from the independence model [114].

2.5 Clave and Clave Direction, with a note on labels

“The brain represents all music and all other aspects of the world in terms of mental[,] or neural[,] codes.” Levitin, [115]

The concept of clave is wrought with controversies. The use of a particular label can offend musicians, start heated arguments, or just cause unnecessary misunderstandings. In the interest of preventing such misunderstandings, a digression is made here to discuss labels in general, to argue that a literal interpretation of a label is

not mandatory. (It is the author's hope that the reader is not offended at the simple analogy employed below. It was chosen because the author has encountered much resistance, from some musicians, to the idea that clave labels may be anything but literal.)

Labels are names given to complex concepts, whether they be musical, physical, mathematical, social, or even for people and things. Many names (which may be thought of as labels for people) in the English language have origins that have mostly been forgotten. For example, *Erica* originally meant ruler or queen; *Michelle* meant god-like; *George* meant farmer, or tiller of the earth. Today, however, when we meet people with those names, we do not expect a *Michelle* to be more religious than an *Amy*, or a *George* to be more pastoral than a *Michael*. Similarly, we do not expect a *Jesús* or *Mehmet* to be more holy or prophetic than a *Guillermo* or *Gökhan*.

Likewise, the labels *3-2* and *forward* for a category of rhythm patterns are just labels, not necessarily with a literal meaning involving the number 2 and the number 3, or any direction of motion.

However, much like names for people, some labels *do* have a certain history of meanings and transformations. *3-2* originally had much to do with the numbers 2 and 3, but in its journeys from Cuba to The United States (and to a lesser extent to Brazil and elsewhere), this label has transformed into a code name for a complex, nuanced concept. Because of its history and the transformations it went through, some practitioners of these musics prefer to keep the label, and some strongly prefer to avoid

it. An example of the latter is discussed in the appendix on Statistics, in Section D.3. Either way, the label does not correspond to the underlying concept literally.

When applied to Brazilian music, the label 3-2 (or *forward*) indicates a relative preference for offbeatness⁶³ (at the level of 8th versus 16th notes) to occur near the beginning and end of bar-long phrases in 4/4 time⁶⁴, and a relative avoidance of offbeatness in the middle portion of such phrases. (It may be better to think of this specifically as “being in *partido-alto*,” as opposed to being in *clave*.)

The terms ‘offbeatness’ and ‘relative’ are both important to understanding *clave* direction, but while the significance of using ‘offbeatness’ instead of syncopation is rather esoteric and intellectual, the significance of using the term ‘relative’ is central to the point.

When one says that 3-2 indicates a *relative* preference for offbeats at the beginning and end of the phrase, it means that the whole phrase must be known before a decision can be made as to *clave* direction. There are no schemata shorter than the entire phrase which can determine without a doubt the *clave* direction of a pattern. All the bits and pieces that make up a pattern must be considered and compared before *clave* direction can be identified. Even Jorge Alabê waited for the experimenter to complete each pattern before deciding whether a pattern played over the 3-2 samba

⁶³ This is only one aspect of *clave* in Afro-Brazilian music. Another aspect has to do with phrasing in a larger context. This latter aspect (in this author’s work tagged *okele* [112, p. 46]—interpreted here as the phrasing aspect of *clave*—as opposed to *chave*—the rhythmic-structural aspect of *clave*) is not discussed here in the interest of brevity and clarity.

⁶⁴ Our “bar-long phrase” is Ekwueme’s *okele*, Danielsen’s *basic unit*, and Arom’s *isoperiod*.

background was crossed or not. Thus we see that even the principal Brazilian samba master in the US cannot (will not) tell the clave direction of a pattern before hearing the entire pattern. That is the nature of clave, at least as developed and interpreted in *samba carioca* (Rio-style samba).

As an example, consider the pattern 10010010. This pattern of eight onsets can start in nine possible places in the 16-onset cyclical phrase (*okele* [116, p. 46]). If it starts in the first position, we are likely to think any resulting pattern is in the 3-2 direction. But the information in that schema is inconclusive because there are traditional samba patterns with this schema in the first position that are in opposite clave directions from one another: 1001 | 0010 | 0010 | 0100 and 1001 | 0010 | 0101 | 0010, the former being 3-2 and the latter being 2-3. It is *by comparison with the rest of the pattern*, not by considering a brief portion by itself, that we see (or hear) this. What follows are two systematic ways to identify clave direction for these two patterns:

1) Intra-pattern Elimination of Onsets

Let's assign subdivisions to the second pattern:

```
1e&a2e&a3e&a4e&a
1001001001010010
oooo          oooo
      iiiiiiiii
```

The letters *o* and *i* stand for “outside” and “inside,” respectively. If there is a pair of subdivision position where one is from the inside and one is from the outside, we can eliminate those onsets without changing the clave-direction of the pattern. This is true for the *a* of 1 and the *a* of 3, so we get:

```
1e&a2e&a3e&a4e&a
1000001001000010
oooo          oooo
      iiiiiiiii
```

We can do the same thing with the *&* of 2 and the *&* of 4:

```
1e&a2e&a3e&a4e&a
1000000001000000
oooo          oooo
      iiiiiiiii
```

We have two onsets left. This new pattern represents exactly the same clave direction as the original pattern. We find the greater offbeatness (in fact, the only remaining offbeat onset) around the middle, on the inside, so we label this pattern 2-3.

2) Inter-pattern Elimination of Onsets

The two patterns to be used for example for consideration for comparing clave-direction compatibility of two patterns are:

```
1e&a2e&a3e&a4e&a
1001001001010010
oooo          oooo
      iiiiiiii
```

and

```
1e&a2e&a3e&a4e&a
1001001000100100
oooo          oooo
      iiiiiiii
```

Let's eliminate the onsets that are in common to both patterns:

```
1e&a2e&a3e&a4e&a
0000000001010010
oooo          oooo
      iiiiiiii
```

and

```
1e&a2e&a3e&a4e&a
0000000000100100
oooo          oooo
      iiiiiiii
```

Do the resulting patterns (with the remaining onsets) carry the same clave direction as one another? In the first pattern, the offbeats are concentrated on the inside, so it is 2-3. In the second pattern, the only offbeat onset is on the outside, and the only on-beat onset is on the inside, so it is 3-2.

These were simple examples. Not only are there cases when there is not such a simple solution, but also, the “inside” and the “outside” are not always clean-cut. Depending on the type of pattern, the boundaries may overlap.

What’s more, the relativity of offbeatness in the present interpretation of the clave direction in samba is *not* whether one would or would not play a certain phrase at a certain point. What is meant is that any such phrase interacts with other phrases around it to form some rhythmic motion, and if such motion or its resolution coincides with the motion and resolution implied by all the other clave-defining ostinati in that piece, then that phrase is in clave, no matter how atypical it may sound in a given style.

The classification of patterns for the training data into clave-direction categories under the various teacher models is based on the following set of criteria, which were arrived at as a result of many years of musical investigation:

1. The *partido-alto* form
2. Occurrence of an isolated missing downbeat (IMD)
3. Gross relativeness
4. *Tercero* pattern from *Mocidade* (partial *partido-alto*)
5. Fine relativeness: relative strengths of schemata and the “algebraic” cancellation of onsets across schema bridges (first to third, second to fourth)

6. Local match with *bossa* and *teleco-teco* (zipper/bookends)
7. African clave sense (antecedent/consequent relationship)
8. “Hanging” onset preceding a missing downbeat (not a full isolation)
9. Direct template-matching to *bossa* clave (when no other criteria apply, such as in the pattern 1111 | 1111 | 1111 | 1100)

These criteria are not absolute. (And perhaps they need to be explained in detail.) Patterns can be arranged such that phrases go across these “tactus boundaries” (given by vertical lines in the binary notation) such that the overall offbeatness or onbeatness of each phrase (relative to one another) is more important than the first-, second-, third-, and fourth-position 4-bit schemata. The “|” symbol is what separates the first, second, third and fourth schema positions in the notation employed throughout this dissertation.

The differences between the firm and strict contexts come down to a few ideas. In the firm context, *partido-alto* is not the absolute ruler. In the strict context, any violation of the P-A (*partido-alto*) form leads to a category 0 (incoherent) output.

Also, should the first two criteria not be thoroughly violated, and the third and fourth are pulling in opposite directions, the relateness of schemata (see below) can be used as tiebreaker, but only in the firm context.

The following example demonstrates what is meant by relateness, using the schema 1101 (“up”) and the schema 1010 (“down”).

1101 1010 1010 1101	very strong 3-2 (<i>partido-alto</i>)
1101 1010 1010 1010	average 3-2 (<i>ijexá</i>)

The latter pattern (*ijexá*) only follows the *partido-alto* three-quarters of the way, but still adheres to the overall *partido-alto* clave direction, as indicated by the following intra-pattern elimination: 1101|~~0000~~|1010|~~0000~~, which leaves 1101 (more offbeat) on the “outside” and 1010 (more onbeat) on the “inside.”

Similarly, the schema 1110 (identified as a “down” indicator in Appendix E) can serve opposing functions, depending on what other onsets are present: It is relatively more onbeat (“down”) than 0110, but relatively more offbeat (“up”) than 1010.

Other examples, presented in a more musician-friendly fashion, can be found in Appendix A.

2.6 Information Theory, Reconstructability Analysis, Multilayer Perceptrons, and the Clave Concept: The Connection

In the psychology experiment described in Appendix B, if a 0 or 1 were somehow stored under each box, the 16 boxes would correspond to the 16 bits of the onset vector. With no boxes lifted, the number of alternatives for the vector stored is $2^{16} = 65536$. This corresponds to 16 bits of potential, or systemic, information. The action of lifting a single box transmits a message that carries a certain amount of Information. It is expected that lifting different boxes will reduce Uncertainty by different amounts toward determining the category of the vector. For example, if the bits (boxes) were labeled 0 through 15 in typical engineering fashion, the musician’s intuition of the author suggests that lifting box 1 should reduce Uncertainty regarding the classification

by somewhat more than lifting box 0, but equal to lifting box 9. Also, lifting boxes 0, 1, and 15 may (state-dependently) reduce Uncertainty more than lifting boxes 2, 3, and 4.

While this much is generally understood by well-versed performers of Afro-Brazilian music, the contributions of lifting other boxes and the interaction effects⁶⁵ between them are not explicitly known, and require the use of information-theoretic (or other forms of) mathematical modeling.

In Information Theory, meaning requires high Uncertainty and high Constraint at the same time. High Uncertainty without Constraint is randomness; high Constraint without Uncertainty is determinism (trivial, where a look-up table may suffice.)

Clave (even if interpreted slightly differently in each idiom) appears to be a rule-based system of hierarchical interactions where clave *direction* is a relative indicator of music-informational *offbeatness* [54]—a Music Information Retrieval term that avoids the cultural baggage of the word *syncopation*). Offbeatness and its distinction from syncopation are explained in Section 1.9.5.

The upshot of this is that the degree of *belonging* to a particular clave direction is obvious to the initiated (*for certain patterns*—attack-point representations of rhythm), but many remaining relationships between clave direction and attack-point patterns are not explicitly known and have not been studied.

⁶⁵ Such as the effects of other boxes (bits) when both bits 1 and 9 are high.

It is a well-known strength of Neural Networks that an artificial neural network can discover and capture effects that are difficult for conscious human analysis to explicitly describe. This is the reason for using clave direction to help improve design practices in the Neural Networks field, and to use neural nets to expand our understanding of clave direction—a sort of bootstrap method where the incomplete function of well-known and well-understood pattern–direction relationships are used to train neural nets, which in turn provide new pattern–direction relationships that augment the experts’ understanding of clave (as practiced in the Afro-Brazilian traditional idiom). Information-theoretic modeling is used to aid both processes by trading off complexity and generalization to identify models that capture Constraint that explains observed relationships and leave enough unexplained to avoid overfitting.

2.7 Researcher’s Background and Qualifications

The author’s relevant academic background is summarized below. For musical background, see Appendix I.

2.7.1 Academic Background as Student

Courses taken in the fields relevant to the present research include Neural Networks I/II, Machine Learning, Statistical Signal Processing, Genetic Algorithms, Discrete Multivariate Modeling I/II, Linear Systems Analysis I/II, Probability and Statistics, Digital Signal Processing, Statistical Communications Theory, and Adaptive Signal Processing (OSU through PSU).

2.7.2 Academic Background as Instructor

Courses taught in relevant fields include Signals and Systems I/II (Electrical & Computer Engineering, PSU, 3 times), Knowledge, Rationality & Understanding (University Studies, PSU, 6 times), Cross-Cultural Rhythm (World Dance Office, PSU), Sound Synthesis (Whitman College Music Department, 4 times in 3 terms), Circuits III (Laplace; OIT, 3 times), and Communication Systems (OIT).

CHAPTER III. LITERATURE REVIEW

The literature review conducted for the proposal for this dissertation (on Music Information Retrieval and Digital Signal Processing) is included in Appendix C. The purpose of that analysis was to establish the feasibility of processing audio so as to represent a series of musical onsets with a binary vector, and also to show that this technology already exists (albeit requiring some modification to perform as needed here; cf. Section 5.3), and is not necessary to duplicate as part of the present research.

The core literature review concerns neural-net prestructuring and optimization. Also examined was the technical literature for several other topics: Information Theory, model selection and evaluation, clave, and common issues in hypothesis testing.

Information Theory, which is crucial to the main research thrust, is explained in Appendix B. Model selection and evaluation (along with the associated topics of network design and data representation) were covered in the Introduction (Section 1.14) and Background (Section 2.2).

The clave-direction concept was introduced in Chapter I, Section 1.17, expanded on in Sections 2.5 and 5.2, and receives a full treatment⁶⁶ in Appendix A.

The problems associated with and techniques for hypothesis testing are profound enough to require their own separate appendix so that the issues can be dissected without a lengthy digression from the main point of the dissertation.

⁶⁶ This is a full treatment of clave direction, but not of all aspects of clave.

Consequently, this portion of the literature research is explained in Appendix D, with some introductory discussion in Section 4.5.

3.1 Neural-Network Prestructuring

The research into prestructured, biased, or modular neural networks and their generalization performance has a 40-year history. This section traces the last 20 years of prestructuring research, starting with highly constrained experiments in which modularization was examined in the context of completely specified I/O mappings, and incrementally moving to less and less constrained cases.

The steps for solving the prestructuring/generalization problem are:

1. Finding or developing techniques for dividing a task into subunits [60, p.1; 61, p. 1];
2. Developing principles for the results of such analysis to be used in prestructuring NNs into modules [60, p. 1; 61, p. 1];
3. Transforming training data for the complete problem into training data for subtasks without losing information (Constraint) necessary for the solution of the complete problem [60, p. 1; 61, pp. 1–2];
4. Determining proper means of combining the trained modules (if necessary, with additional training of the inter-module connections) [60, p. 1; 61., p. 2];
5. Testing the applicability of prestructured NNs to more general problem contexts:
6. Investigating generalization when the full mapping is not known in advance (which is the proper use of neural networks) [60, p. 1];

7. Investigating generalization of a NN trained on a binary representation to an analog, or, at least, non-binary quantized pseudo-analog representation [117].

The present research focused on steps 3, 4, and 5, and involved replication and verification of steps 1 and 2 as well.

3.2 Literature Review for Neural-Network Prestructuring

In their 1992 IJCNN paper, Lendaris and Todd used a fully known problem context (wherein table lookup would otherwise suffice) to explore whether a properly modularized network could achieve the same complete mapping as a fully connected network trained to 100% accuracy. Their approach utilized existing literature on knowledge representation using conceptual graphs [118, 119] for the input domain, and the graph theoretic *Connection Matrix* and *Reachability Matrix* methods [120, 121] for the prestructuring and interconnection of the modules.

A 25-node lattice was modularized into 16 subunits trained separately. They reported that, after interconnecting the modules, “training to the 100% success level was typically achieved within 30 epochs (of the n^2+2n training exemplars)” [60, p. 5]. (This constitutes a demonstration of prestructuring step 1 of Section 3.1.)

For a different type of problem, Lendaris, Zwick, and Mathia showed that appropriately modularized neural nets generalize better than fully connected neural nets for decomposable functions [61]. (It is interesting from the standpoint of intuition that modularized neural nets do not learn non-decomposable mappings.)

The prestructuring was based on identifying Constraints among the input and output variables describing the problem to be solved, and using such Constraints—represented in terms of the General Systems Methodology (GSM) lattice of structures—to graphically translate inter-variable relations into NN modules [61, p. 2].

Multiple Boolean functions were employed in the learning experiments. These included three-input and five-input, decomposable and non-decomposable functions, each with one Boolean output. The experiments demonstrated both the reliable success of modularized networks with decomposable functions, and the expected lower degree of success (generalization) of fully connected networks when dealing with decomposable functions.

Furthermore, generalization was investigated by selecting a representative data set using knowledge of the problem domain, and containing only half of the full mapping. While the fully connected NN generalized to only 55% of test data, the prestructured NN was shown to generalize to 100% accuracy.

These experiments focused on steps 2, 4, and 5a; the information required by steps 1 and 3 were part of the problem setup. In other words, since “functions were constructed whose decompositions were known *a priori*” [61, p. 6], the models used did not need to be determined using such model-search procedures as Principle-Component Analysis (PCA), Reconstructability Analysis (RA), Extended Dependency Analysis (EDA), Discriminant-Function Analysis (DFA), Factor Analysis, or Cluster Analysis. (For the research put forward in this proposal, RA is intended for modeling

the underlying problem based on training data. Details on the proposed use of RA, including issues of RA generalization in the presence of prohibitive variable-cardinality—the curse of dimensionality—are discussed in Section 4.6.5.)

Subsequent papers co-authored by Lendaris on prestructuring and generalization in 1994 extended the investigation of neural-net theory via prestructuring by relating experimental findings to complexity measures and the concept of a *performance subset* [22, 62, 63]. Performance subset is defined as “that set of functions the [network] is capable of performing as the element parameters range over their possibilities with the structure remaining fixed” [122, p. 500]. To paraphrase, a performance subset is the set of mappings that could possibly be achieved using all possible weight combinations by the neural network for a given connection scheme and activation function⁶⁷.

As part of their research for the GM Defense Research Laboratories, Lendaris and Stanley conjectured in 1965 that “if an ANN successfully learns a training set, then the smaller the ANN's performance [subset], the better will be its generalization” [122, p. 501]. This conjecture has implications for prestructuring based on the concepts of performance subset (PS) and set of all possible mappings (SAPM), described below.

For a Boolean mapping with n inputs and one output, there are 2^{2^n} possible mappings. This is known as the set of all possible mappings (SAPM) [62, p.1]. By virtue

⁶⁷ These two aspects of a neural network, along with a third, the training algorithm, are jointly referred to as the *paradigm*.

of the fact that all *possible* mappings from n variables to one variable are considered, the SAPM includes potentially nonsensical mappings, in addition to useful and meaningful ones. The SAPMs for two hypothetical cases are shown in Figure 23.

The performance subset (PS) is the set of all mappings that can be performed by a specific neural-net structure [above, and 22, pp. 464–465]. In Figure 23a, the performance subset is smaller than that of Figure 23b.

One possible way in which an instance of a neural network (such as Figure 23a) can have a performance subset that is a proper subset of the performance subset of another instance with the same I/O scheme, structure, and element type (such as Figure 23b) is if the interconnections in the former are limited by design to a subset of the interconnections of the latter. In other words, prestructured neural nets have smaller performance subsets [62, p. 11].

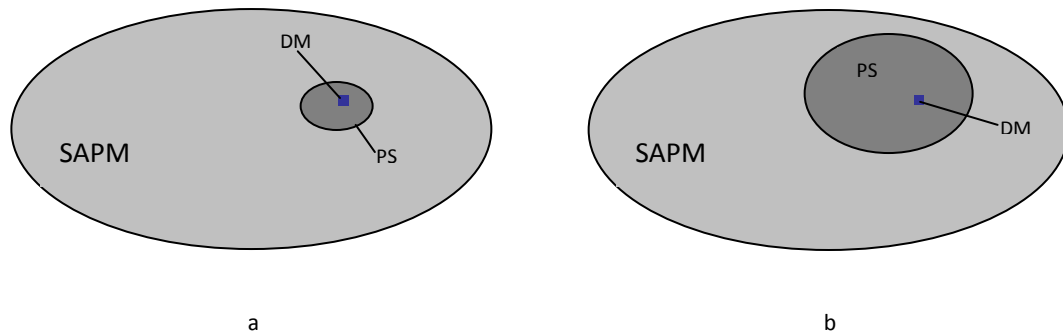


Figure 23: The set of all possible mappings (SAPM, light grey oval), the performance subset (PS, dark grey oval), and the desired mapping (DM, dark dot) for two hypothetical neural networks

Let us assume that the former is a prestructured feedforward neural net, while the latter is fully connected, as illustrated in Figure 24a, and Figure 24b/c, respectively. It is easy to see that the fully connected network of Figure 24b will be capable of realizing a greater number of mappings, including all the mappings that can be realized by the modular network.

At the same time, interpreting Figure 23 as a Venn diagram where areas correspond to probabilities, one can see that the probability of realizing a mapping other than the desired mapping (DM) is greater with the bigger performance subset. For a detailed explanation of why this is the case, see [122, p. 501].

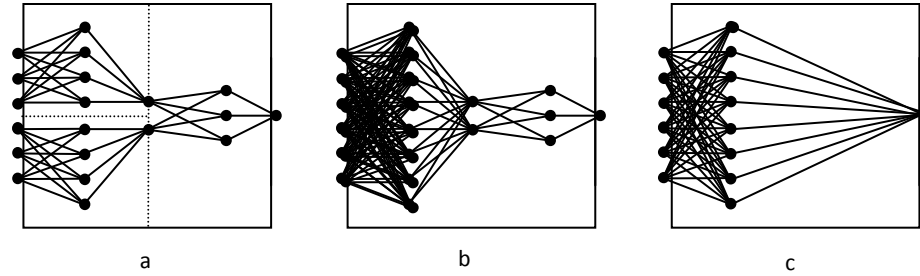


Figure 24: Modular (a), and fully connected (b, c) neural networks

3.3 Current Practice in Neural-Network Optimization

Both modularization and the information-theoretic selection of features and training vectors have been proposed as schemes of generalization improvement for neural networks (along with a variety of evolutionary schemes). The proposed research has the advantage of combining modularization and information-theoretic feature

selection through rigorous data analysis. Furthermore, the proposed method is systematic (a well-defined series of steps and guidelines), not a collection of heuristics like some of the other performance-improvement schemes and alleged methods of optimization.

“Improving Generalization Performance by Information Minimization” by Kamimura, Takagi and Nakanishi [123] is essentially a restatement of the Lendaris-Stanley conjecture about generalization [122, p. 501]. The authors relate network complexity to Information capacity, and that in turn to overfitting. Hence they propose that the network architecture, as measured by Information in the hidden-unit activations, should be as small as possible without greatly deteriorating the level of learning of the training set. The mathematical derivations of Information, Uncertainty and back-propagation are standard, but the proposed function, T , to be minimized for information-theoretic network training appears arbitrary and unfounded. Furthermore, T varies linearly and proportionally with Uncertainty, minimizing which actually maximizes Information. It is the reviewer’s deduction that the authors have confused the terms Uncertainty and Information, and that what they intend to minimize in the hidden layer is actually Uncertainty. In addition, they provide a plot of training-data performance and refer to it as “generalization” (p. 146). The second plot of the same figure does indeed relate to generalization, and if this evidence were shown to be more than anecdotal; i.e., statistically significant, there would be reason to consider the proposed method of Uncertainty reduction among hidden-layer units.

It is important to note that the learning algorithm used is not back-propagation, in the sense that squared error is not the criterion for learning; Information and Uncertainty are.

While the authors conclude that information-theoretic learning leads to better generalization, the plots suggest only one run per method was carried on only one problem. Questions remain as to whether the improved performance was statistically significant and whether it implies similar improvements for non-linguistic test sets. These problems, in addition to the questionable derivation of the objective function (p. 144), leave unresolved issues regarding this technique.

In “Automatic Determination of Optimal Network Topologies Based on Information Theory and Evolution” [124], Ragg, Gutjahr and Sa search for “the most efficient optimization technique” which seems to suggest that optimization itself is a closed issue, and only computational efficiency of optimization is under investigation. This turns out not to be the case for universal neural-net concerns when the authors define optimization as in the following: “relevant information to solve the problem is extracted from the input data in order to reduce the input dimensionality to the most possible (sic) limit.” This is feature selection, which is only one aspect of NN optimization. Nonetheless, the authors show generalization improvement for four very different problems (without addressing statistical significance of the *very* small improvements demonstrated), and also state that the technique described (ENZO (MI)) is available as a free download. This makes it possible to compare generalization

performance upon prestructuring with how ENZO (MI) performs on fully connected networks.

The weakness of this approach seems to be its focus on the input layer. In fact, according to the conclusion, the method of regularization (weight decay) used in the tests described in the Ragg, et al. paper was not an integral part of the ENZO (MI) mechanism.

Another detail of note is that the fully-connected networks had shortcut connections (layer-skipping connections, also known as prior connections when between the input and output layers). This is a wildcard in that it increases complexity, but may also emerge as a form of prestructuring.

In “Hybrid heuristics for optimal design of artificial neural networks” [125] Abraham and Nath define optimal as “smaller, faster and with a better generalization performance,” raising questions like “better than what?” and “smaller by how much?” They also state that “optimal design of an ANN can only be achieved by the adaptive evolution of connection weights, architecture and learning rules.” The claim that this is the *only* possibility is curious, at best.

In any case, the method described is ‘Genetic Annealing’ (GAA), a combination of Genetic Algorithms (GA) and Simulated Annealing (SA), which uses the SA concept of energy to accept or reject GA-generated individuals. Three separate GAA runs are proposed to generate ‘optimal’ connection weights, learning rules (parameterized linear-combination versions of standard backprop), and connection

schemes. The suggested genetic representations of these three aspects appear simple and promising. As in the last paper, the training is not regular backprop gradient descent; this time it is evolutionary recombination, mutation and selection. Back-propagation is, however, allowed to take place in the later stages of evolving and learning connection weights.

It is ironic that the authors are trying to optimize all these neural-net parameters that are commonly determined heuristically, and in order to do so, they have introduced quite a few heuristic GA and SA parameters. It is the reviewer's opinion that this defeats the purpose of lesser reliance on heuristics.

The paper does not discuss in detail, but draws some attention to, other methods, such as Extentron, Upstart, L-Systems, Symbiotic Adaptive Neuro-Evolution (SANE), Cellular Encoding, Fractally Configured Neural Networks, Marker-Based Genetic Coding, and the Graph Generation System. A statistical study may, at some point, be needed to compare all of the existing evolutionary, information-theoretic, and other methods of neural-net design and generalization improvement. It is regrettable that the paper does not show any results.

“A Comparative Study of Neural Network Optimization Techniques” by Ragg, Braun and Landsberg [126] suffers from significant mistakes in English usage at critical points, making it difficult to understand the nature of network connections (“connections between around all layers,” possibly meaning prior connections) and pruning schemes (“prune minor important weights”) described in the paper.

The main argument that is relevant to the work at hand is that network optimization is defined foremost in terms of size, then computation time, then classification error. This is almost the reverse of the priority order to be undertaken in the reviewer's proposed research in that classification error (and its relationship to generalization) is paramount, network size reduction is a consequence of increased generalization, and computation time is not a current concern.

Furthermore, the authors also seem to imply the opposite of the Lendaris-Stanley conjecture (see above and [122, p. 501]), but this may again be due to poor use of language: "the smallest network achieving a learning error below a given error limit has not the best generalization performance in general."

The paper dismisses all so-called constructive techniques (such as CCLN), saying they lead to overfitting in the absence of large training sets. Constructive techniques are those that add units to an existing NN architecture. Destructive techniques, like pruning, remove units from an existing architecture.

The bulk of the proposed improvement seems to reside in connection pruning and weight decay. However, they then state that the weight-decay factor was optimized prior to experimentation (how?) and that the demonstrated improvement is "mainly due to the elimination of redundant input units." The unclear explanations and plots that are too small to make out render the claims in this paper questionable.

Bianchini and Gori [127] introduce the concepts of "Decoupling Networks Assumptions (DNA)" (p. 2) and "pattern-mode learning" (p. 3), which states that

difficult patterns ought to be presented to the network more often than others. The meaning of this statement turns on what is implied by “difficult.” The authors put this word in quotation marks, but neglected to explain what they meant by it. It is the opinion of the reviewer and his research adviser that if difficult patterns are those that are difficult for human experts to classify, this is the exact opposite of what should be done. On the other hand, if the authors used the term “difficult” to mean “close to decision boundaries,” then their claim may be substantiated. This relates to another recommendation for NN training made by Drucker and Le Cun: “In order to generalize from a training set to a test set, it is desirable that small changes in the input space of a pattern do not change the output components” [128]. This is *not true* for the recognition of clave direction, and illustrates the difficulty of the clave-classification problem. In clave direction, a single-bit change in the input can cause a change in the output class (though not always). The “double-backpropagation” method espoused by Drucker and Le Cun is based on this assumption. Thus, at the outset, it appears antithetical to the clave-direction problem. It also introduces additional heuristics, such as the use of noise when dithering for local-optimum avoidance [128, p. 991].

Moreover, the double-backpropagation network is approximately twice the size of the regular network in order to allow for the calculation of the Jacobian—which requires a forward and a backward pass through the network—and the four architectures used in conjunction with double-backpropagation are already prestructured in the sense of receptive fields and selective connections [128, p. 995].

Hence, the benefits of double-backpropagation and the benefits of prestructuring are conflated.

In “Optimal Linear Combinations of Neural Networks” [129], Hashem investigates a related problem: how to combine multiple (differently structured) networks each trained on the same problem. This is potentially useful information for the research proposed in this proposal because information-theoretic modeling offers many mathematical models for prestructuring neural nets given some set of training data. Linear combinations of more than one model present an alternative to selecting a single best model, or making a committee-style decision for model and structure selection. Otherwise, the present research and Hashem’s methods are “orthogonal” in that the former primarily concerns how to design modules while the latter concerns how to combine existing modules.

Hashem defines optimal as minimal mean-squared error and presents four ways to obtain what is referred to as an MSE-OLC (mean-squared error-optimal linear combination). Network learning rates, momentum and weight decay are taken into account, along with topology and connection weights (p. 1). After discussing the theoretical (statistical) estimation of the mean-squared errors for the four linear-combination schemes, Hashem also examines generalization performance for two function estimations. The conclusion is that MSE-OLCs of trained neural nets perform better than the individual nets. While a significantly different concept than prestructuring, the notion of combining trained networks can be useful when modeling suggests multiple prestructuring schemes for a given problem. The so-called

“unconstrained MSE-OLC with a constant term” is recommended for function approximation, but the discussion throughout the paper suggests constrained MSE-OLCs perform better in terms of generalization. The four candidates for linear combinations, however, are given in terms of regression expressions (for function approximation) and need to be adapted to classification problems to be useful.

Another study of modularization and module combining is reported by Lu and Ito [130], who state that any K -class classification problem can be decomposed into a series of two-class problems by ignoring $K-2$ classes at a time, and without any explicit domain knowledge on the part of the designer. While the exploitation of relations in the data is similar to Reconstructability Analysis, the decomposition does not go to extent of making choices among input variables in terms of their contribution to determining the class of an object. Also, the basic example given by Lu and Ito is too clean to be a good model for the clue-direction problem. Nonetheless, the module-combination scheme could be of potential value in further prestructuring a neural net, in this case in terms of the connections from the hidden layer(s) to the output(s). Lu and Ito propose a generalized conjunction/disjunction method of combining the decisions made by different classifiers. Identical outputs from different classifiers for the class under consideration are combined using the MIN function (generalized AND) so that an example is classified into that class only if the classifiers agree. Any outputs indicating other classes are combined using the MAX function (generalized OR) such that examples are easily classified away from the class under consideration, since in the

scheme of Lu and Ito, all classes other than the one under consideration are temporarily lumped together.

There is much more research in neural-net optimization with varying degrees of success demonstrated in the literature and a wide variety of methods (though they all fall under the heading s of statistical, evolutionary, or information-theoretic approaches). Future research is needed to compare RA to the better among these methods.

CHAPTER IV: METHODOLOGY

4.1 Introduction

The approach to the experimental design of the present research owes its current form most significantly to the influence of the following sources:

- Kapur, J. N., and Kesavan, H. K., *Entropy Optimization Principles with Applications*, Boston, MA: Academic Press, 1992. [15]
- Box, G. E. P., Hunter, W. P., and Hunter, J. S., *Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building*, New York, NY: John Wiley & Sons, 1978. [21]
- Hastie, T., Tibshirani, R., and Friedman, J. H., *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, New York, NY: Springer Series in Statistics, 2009. [39]
- Salzberg, S. L., “On Comparing Classifiers: Pitfalls to Avoid and a Recommended Approach,” *Data Mining and Knowledge Discovery*, 1, pp. 317–327, 1997. [33]
- Freedman, K. B., and Bernstein, J., “Current Concepts Review - Sample Size and Statistical Power in Clinical Orthopaedic Research,” *J Bone Joint Surg Am.*, No. 81, pp. 1454–60, 1999. [131]
- Siegfried, T., “Odds Are. It’s Wrong: Science fails to face the shortcomings of statistics,” *Science News – The Magazine of the Society for Science & the Public*, Vol. 177, No. 7, p. 26, 2010. [132]
- Ioannidis, J. P. A., “Why Most Published Research Findings Are False,” *PLoS Medicine*, Vol. 22, No. 8, p. e124, 2005. [133]
- Miller, J., “Statistical significance testing—a panacea for software technology experiments?” *The Journal of Systems and Software*, 73, pp. 183–192, 2004. [134]
- Zucchini, W., “An introduction to model selection,” *Journal of Mathematical Psychology*, 44, pp. 41–61, 2000. [135]
- Prechelt, L., “A quantitative study of experimental evaluations of neural network algorithms: Current research practice,” *Neural Networks*, 9, 1996. [136]
- Flexer, A., “Statistical Evaluation of Neural Network Experiments: Minimum Requirements and Current Practice,” In R. Trappl, ed., *Cybernetics and Systems ’96: Proc. 13th European Meeting on Cybernetics and Systems Res.*, pp. 1005–1008, Austrian Society for Cybernetic Studies, 1996. [137]
- Holte, R., “Very simple classification rules perform well on most commonly used datasets,” *Machine Learning*, Vol. 11, No. 1, pp. 63–90, 1993. [138]
- Tversky, A., and Kahneman, D., “Belief in the law of small numbers,” In D. Kahneman, P. Slovic, and A. Tversky (eds.), *Judgment Under Uncertainty: Heuristics and Biases*, Cambridge University Press, pp. 3–20, 1971. [139]

These books and articles, in combination, gave the author first an appreciation of, and then a cautious skepticism toward the methods of Statistics as they relate to the validity and usefulness of the scientific method. Before delving into the details of the statistical pitfalls of experimental design (in Section 4.5), however, it is necessary to introduce the nature of the data and its representation.

4.2 Notation and Data

4.2.1 *Occam3* Notation

The *Occam3* notation assigns letters to each categorical variable, whether independent (input) or dependent (output).

In order to avoid confusing the letter *i*, the letter *l*, and the number 1 (and other similar sets), the following correspondence was set up in the data files:

ABCDEFGHJKLMNPQRS are the input variables such that they are in the order of 16th-note onsets in a bar. Z is the output variable representing the clavedirection category.

The categorical variable A corresponds to the presence or absence of a qualifying note attack on the first downbeat, a 0 signifying absence, and a 1 signifying presence (positive logic). Similarly, the variable B corresponds to the presence or absence of a qualifying note attack on the 16th-note immediately following beat 1.

Likewise, the variable F corresponds to the presence or absence of a qualifying note attack on the 16th-note immediately following beat 2, and so on.

1	e	&	a	2	e	&	a	3	e	&	a	4	e	&	a
A	B	C	D	E	F	G	H	J	K	M	N	P	Q	R	S

Figure 25: 16th-note subdivisions and the corresponding *Occam3* variable names.

The output (*Z*) representation used for *Occam3* is:

- Neutral (no clave), category 3,
- Reverse/2-3, category 2,
- Forward/3-2: category 1,
- Incoherent (no clave direction): category 0.

4.2.2 Music Notation

The justification for the notational preference (sixteen 16th notes per phrase, whether expressed or rests, constituting one bar of 4/4) is given in the section on attack-point rhythm, the survey of Afro-Cuban and Afro-Brazilian music-instruction literature in Appendix J, in combination with Danielsen's synthesis of Kwabena Nketia and Arom [53] in Appendix C.

4.3 Nature of the Data

Attack-point rhythm (limited to 4/4 time in the present research), quantized to a 16th-note tatum (i.e., no tuplets, no swing), is a defensible abstraction and a feasible task in three ways:

1. Psychoacoustically in that at least some musicians and listeners conceptualize sequences of sound by recreating them in a spatial representation [140];
2. Musically in that it is a shared representation in the Afro-Brazilian and Afro-Cuban music community in the US and parts of Europe and Asia (defined as the faculty and students of California Brazil Camp⁶⁸ and Humboldt State University's Afro-Cuban camp⁶⁹); and
3. Technologically as shown by the survey of Music Information Retrieval (MIR) in the literature review to the proposal, included here as Appendix C, in that it can be obtained from audio or MIDI through existing or rapidly developing MIR DSP techniques.

The result is a total of $2^{16} = 65536$ possible rhythm patterns. The complete set of these 65536 patterns was created using a C program in 2006, and scrambled using a high-quality random-number generator, also in C.

Each 16-bit attack-point onset vector indicates (in each bit, or position) the presence or absence of a significant note onset in an appropriate time window.

⁶⁸ <http://www.calbrazilcamp.com/>

⁶⁹ <http://www.humboldt.edu/afrocuban/>

Significant onset implies a note attack whose slope and final value within the window surpass an experimentally determined, musically meaningful threshold.

The 16 bits described above make up the “input” portion (independent variables) of each datum. The output, or dependent variable, can be expressed in a variety of ways. Some of these variations are requirements of the technology being used—for example, *Occam3* can only have one output field, so no multi-bit binary coding—and some are necessitated by the nature of neural-net experimentation. The latter concern arises because the success of a neural network depends as much on the representation of data (inputs and outputs) as it does on the structure, learning algorithm, processing unit type(s), and parameters.

The first three output encodings used for neural nets during the exploratory phase of the research were true binary, pseudo-one-up (“GGL encoding”) and true one-up, shown below in the right column, listed left to right:

- | | |
|--------------------------|---------------|
| • Neutral (no clave): | 11, 100, 1000 |
| • Reverse/2-3: | 10, 010, 0100 |
| • Forward/3-2: | 01, 001, 0010 |
| • Incoherent (no clave): | 00, 000, 0001 |

The encoding used in the final experimental thrust is the one-output-at-a-time encoding, which is 1 for the direction under consideration, and 0 for all three remaining categories.

This encoding would require all experiments to be done triple the number of times as with the other encodings. The final product would then have three

corresponding neural networks, one for the forward direction (forward 1, all other categories 0), one for the reverse direction (reverse 1, all others 0), and an additional network to combine the first two networks' outputs for a final decision on clave direction. However, since any one of the aforementioned networks is sufficient to investigate prestructuring and generalization properties, the requirements of the present research into Neural Networks and prestructuring can be completed with one set of one-output-at-a-time networks.

An additional aspect of the data (which was not encoded into the vector representation, but was noted with each I/O pair to inform design practices and interpretation of results) is the distribution of membership degrees in each of the four output categories. Though the descriptive names for membership degrees vary in accordance with the musical differences among the clave-direction categories, the four categories essentially have the same five levels: strong, very strong, average, weak and very weak. During the act of iterative classification (2007–2009), these labels arose naturally through the noting of observations as to the musical (rhythmic, clave-direction) impression from each pattern. The particular examples listed in this section are for clave-direction categories 1 and 2; similar levels with different wording were used to reflect corresponding insights into categories 0 and 3 (such as 'fully neutral', 'weak neutral', 'good zero' and 'average zero').

After classification and annotation, these labels were grouped into three larger groups that are more practical: STRONG to comprise 'strong', 'very strong', 'good' and

‘fully’; AVE to mean the average level for each category; and WEAK to comprise ‘weak’, ‘very weak’, and several corresponding labels for categories 0 and 3.

Although all classifications and annotations have been thoroughly and repeatedly reviewed, the concepts remain far from crisp (as musical concepts expressed in such a wholesale manner are bound to). Hence, the five degrees are further grouped into three sets by the researcher: STRONG, AVERAGE, and WEAK (where STRONG includes what was annotated as both ‘strong’ and ‘very strong’, and similarly with WEAK). These three labels are clearly discernible by the author, both as to their meaning in terms of clave-direction in the *partido-alto* paradigm and in terms of telling the three labels apart when given a particular vector. That is, while there may be some ambiguity between ‘weak’ and ‘very weak’, there is no ambiguity between WEAK and AVERAGE.

The STRONG group includes the clearest examples of *partido-alto* itself, and of *partido-alto*-based clave direction (such as 0101 | 0010 | 1010 | 1001), along with the most unmistakably neutral or incoherent patterns (such as 1111 | 1111 | 1111 | 1111 or 1000 | 1000 | 1000 | 1000 for neutral, and 1111 | 1010 | 0101 | 1000 or 0111 | 0110 | 0101 | 1111 for incoherent).

The AVE group includes the most informative examples of clave direction (as well as the less informative examples of its lack), such as 1111 | 0110 | 1111 | 0101, 1110 | 0011 | 1110 | 0111, 1101 | 1011 | 1101 | 0100, or 1100 | 0010 | 1110 | 0010 (average members of the forward category) and 1111 | 1111 | 1111 | 1110, 1110 | 1101 | 1110 | 1110,

1111|0010|0111|1010, or 1110|1100|1000|1000 (average members of the reverse category).

The WEAK group constitutes the most stringent test of the notion of clave direction, where different indicators point in opposite directions (such as an isolated missing downbeat strongly suggesting forward, but the rest of the pattern closely following the reverse direction). An example of this would be 0100|0101|0100|1011. This example brings up another issue discovered and documented during the clave-research and classification processes, and which eventually informed the experimental design. Under different circumstances (for instance, performance versus teaching) the same teacher may classify the same questionable or difficult pattern as acceptable in a given clave direction or unacceptable. It has been observed not only that the strictness levels of teachers vary with their background (those with born-and-raised credentials can afford to be more lenient), but also a particular teacher's opinion may change based on context as well. (Hence clave direction as an ill-defined problem, as discussed in Chapter I.) To address these possibilities, and more importantly, to inform the process of classification based on these possibilities, the present author has actually classified the 10,784 patterns into *three* classifications based on possible teacher attitudes to clave: strict, firm, and lenient.

The strict teacher classifies into incoherent (category 0) *any* pattern that goes against the flow of the *partido-alto* prototype (which is 0101|1010|1010|0101 for

forward, and 1010|0101|0101|1010 for reverse)⁷⁰. As a simple example, 0101|0101|1010|1010 would break the *partido-alto* pattern in either direction, while 1011|0111|0110|1011 would still be well within the reverse direction's range.

The strict teacher would also reject (classify as incoherent) *any* pattern with opposing tendencies according to any two or more of the paramount criteria.

The firm teacher, while taking such inconsistencies into account, would give greater leverage to “relativeness”; i.e., consider how strongly each indicator points in each direction, and also take into account the overall sound of the pattern, and can thus accept some patterns rejected by the strict teacher, though not many.

Another important difference is between the strong and firm teachers and the lenient teacher. Since clave direction can also depend on the role a pattern plays in the music (is it a riff, *desenho*, or ostinato—something repeated, or is it an entrance or a one-time lick in improvisation?) as well as the instrument it is played on (the bass boom of the *surdo* versus the crisp strumming of the *cavaco*), the strict and firm teachers only accept a pattern as being in a clave direction if it would work in both cases on any instrument, whereas the lenient teacher can call it a weak neutral if it would primarily work in some circumstances as an ostinato, and be temporarily tolerable on other instruments.

⁷⁰ These prototype rhythms are the most recognizable of the standard samba patterns, frequently played on the *pandeiro* in the style known as *pagode*. For a brief and clear description of what constitutes a *prototype*, see Allman's discussion of categorical hierarchy [14, p. 167].

The engineering work in this dissertation is based solely on the firm-teacher model, but the recognition and development of the other models was necessary for this model to be internally consistent.

The full data set (before holdout removal) consists of 10784 vectors. Table 6 shows the breakdown of these vectors into categories and gross membership degrees.

Table 6: Breakdown of the full data set into clave-direction categories and gross membership degrees.

	STRONG	AVE	WEAK	
0	1443	342	0	<u>1785</u>
1	2118	1074	1105	<u>4297</u>
2	2119	1020	1150	<u>4289</u>
3	244	61	108	<u>413</u>
	<u>5924</u>	<u>2497</u>	<u>2363</u>	10784

The entropy of the data set (10784) prior to classification is given in (44).

$$\log_2(10784) = 13.4 \quad (44)$$

After classification into the four categories 0, 1, 2, and 3 (Table 6), the new Entropy value is given by (45) based on (27):

$$\begin{aligned}
H(A) &= \sum_{a \in A} \frac{n_a}{n} \left(-\log_2 \left(\frac{n_a}{n} \right) \right) \\
&= \frac{1785}{10784} \left(-\log_2 \left(\frac{1785}{10784} \right) \right) + \frac{4297}{10784} \left(-\log_2 \left(\frac{4297}{10784} \right) \right) \\
&\quad + \frac{4289}{10784} \left(-\log_2 \left(\frac{4289}{10784} \right) \right) + \frac{413}{10784} \left(-\log_2 \left(\frac{413}{10784} \right) \right) \\
&= 0.429515 + 0.528953 + 0.529037 + 0.18025 = 1.66776
\end{aligned}
\tag{45}$$

This value is very close to that obtained after classifying all 65536 instances into four arbitrary categories (1.799, Section 2.4.2). This makes sense because the distribution of the vectors into the four sets (in either instance) is relatively even (compared to one set having all or almost all the members, for instance, or one set having none or close to none). In the case of the fully even distribution, Entropy would be equal to 2, so any reasonably even distribution will have Entropy close to 2. If all vectors were classified into the same category, Entropy would be 0, and Uncertainty reduction would be 16. Hence, we see that the Entropy value alone says nothing about the type of classification done. Arbitrarily assigning vectors to output categories would also serve to reduce Entropy. The only conclusion available from this calculation is that Uncertainty has been reduced by more than it would have been if vectors had been distributed evenly into the four clave-direction categories. However, 1.66776 is much closer to 2 (maximum entropy) than it is to 0 (maximum reduction of uncertainty), meaning the sole act of classifying has hardly brought us closer to a practical

understanding of what category a particular vector should belong to. Entropy serves merely as a descriptor of spread. (See Section 2.4.2.)

4.4 The Determination of Data Splits

10784 (the total number of I/O pairs) is 16.5% of the problem space of 65536. Of these data, the design set and out-of-sample holdout can be split in two ways. One is the *random-holdout* split; the other is the *weak-holdout* split. These correspond to two training strategies for neural nets.

In the former, the holdout for final testing and assessment of the networks' generalization performance is based on a statistically sound random selection. In the latter, the networks are trained on a better selection of examples for learning (based on the idea that biological neural networks learn best from good examples) and tested for final performance assessment on "weak" (difficult) examples.

The former split is 8631 design vectors and 2154 out-of-sample vectors. The latter split is 8442 design vectors ("good" examples) and 2369 "difficult" test vectors.

These data are annotated as to the degree of the membership of each vector in its clave-direction category. These degrees of membership arose naturally in the course of classification (2006 through 2010) and fell into the five levels of membership described above in Section 4.2.1.

The design data in the random-holdout split consists of 4741 STRONG examples (55% of the design data) and 2009 AVERAGE examples. These together make 78% of the design data for this split in the case of the "average and up" training/validation set.

The design data in the weak-holdout split consists of 5924 STRONG examples (70% of the design data—naturally a higher number than in the other split because all STRONG examples in the design set are included) and 2497 AVERAGE examples. These, by design, make up 100% of the design data for this split.

The breakdown of the data into sets, categories, and membership degrees is shown above in Table 6.

4.5 Experiment-Design Rationale

The quality of any body of research is a function of the quality of its experimental design. Experiments, along with the meta-science of interpreting their results (Statistics), are necessary in the study of complex systems for two reasons.

One is that most complex systems involve multiple *interacting* factors that influence their organization, behavior and outcomes. From the dismissal of the doctrine of unique etiology in medicine⁷¹ [141], up to the development of the sciences of complexity, systems and emergence, the notion that complex systems (such as the environment, social networks, music, or health) have multiple interacting components has gained a central position in our efforts to understand physical, biological and cultural phenomena, which collectively guide Computational Intelligence technology.

⁷¹ It is not unusual to equate the doctrine of unique etiology with the germ theory of disease (cf. [131, pp. 141–153], and the end of said doctrine with holistic, naturopathic approaches to medical care, but neither is entirely the case. While there are superficial relationships between each of these associated hypotheses, the connection between the germ theory of disease and the acceptance of multiple etiology through the scientific method is even stronger: The immune system *and* the placebo effect are more widely, thoroughly and rigorously studied by the mainstream scientific community (the *scientific* scientific community) than the pseudoscience community, according to [142] written by the former director of research at the University of Maryland’s Complementary Medicine Program, an insider to so-called alternative medicine.

The other reason for the importance of careful and informed experimental design and statistical analysis is the infeasibility of observing and recording all variables (factors) for all time in all agents of interest. Whether the issue at hand is projecting vote counts at an election, automated recognition of a song over a cell-phone connection, or curing cancer, modeling the systems in question through some form of sampling of relevant factors is necessary to advance our understanding within practical budgets and time frames.

So, we sample. And, thanks to the work of hundreds or thousands of statisticians and mathematicians (as well as the worldwide network of scientists, and the check and balances of the scientific method), we have the tools to pursue understanding and knowledge to the best of our collective epistemic, mathematical and technological means.

The tools, however, are many and complicated, and often misused. Does a particular research thrust require hypothesis testing with significance, or confidence intervals, both, or neither? When are assumptions of independence valid? How about assumptions of underlying distributions? (Are they really Gaussian so often?) Have variances among groups been measured to aid in test selection? How does one choose among the many (and conditionally but not completely equivalent) techniques of statistical estimation, such as AIC, BIC, CIC, DIC, FIC, NIC, RIC, TIC⁷² (some of the

⁷² AIC is alternately known as the Akaike Information Criterion or “An Information Criterion.” BIC stands for Bayes Information Criterion, but is also known as the Schwartz Information Criterion. CIC appears to mean “Curvature Information Criterion.” *D* is for “deviance.” FIC is the Focussed (sic) Information Criterion. The *N* in NIC stands for “network” where an information criterion is used to

many penalty schemes for model selection), or the many forms of cross-validation, bootstrapping, jackknifing, the lasso and other techniques (each of which may be equivalent to another under certain circumstances of model types and data-set sizes and not others)? How to interpret statistical significance when at least three conceptual problems are widely acknowledged in relation to that notion⁷³? For that matter, how to choose a target significance level for a study? Whether to block or not? When to use boosting or bagging? How to apportion data between training and test sets, and to choose the type of sampling that is most appropriate (random, stratified, or other)?

For both statistical validity and practical usefulness, a thorough investigation of the Statistics literature is needed prior to the design of any experiment, be it the design and comparison of neural-net classifiers or an epidemiological study. The following subsections introduce the present inquiry into aspects of Statistics relevant to the design of such CI experiments as needed for the present research: factorial design, statistical power, statistical significance, the choice of appropriate hypothesis tests, the design of controls, assumptions of independence and underlying distributions, and various approaches to model selection and assessment.

4.5.1 Factorial Designs

Factorial experimental designs are those situations in which multiple factors need to be investigated. Any design problem in Neural Networks would be an example.

determine the number of hidden-layer elements in a neural net, *R* is for “robustified,” and the *T* in TIC is for Takeuchi, who developed one of the many variations of AIC.

⁷³ For details on those problems, see Section 4.3.2.

Even if we consider only the numerical parameters, the designer must plot a course for experimenting with and determining network momentum, derivative offset, step size, the number of hidden layers, the number of elements per layer, and other parameters. Even non-numerical aspects of the design, like the choice of learning algorithm or activation function, input and output encodings, and learning schedules appear as experimental variables that are interdependent in the sense that any change in one is likely to affect future behavior due to changes in the others, and that there is no order in which these parameters have to be determined or decisions made.

The (fictional) example of the helmet type and auto speed regarding collision fatalities (Section 2.4.1) shows how in a factorial study—which the helmet data, ethically, could not have come from, but nonetheless has the same tabulated form—outputs may vary in their direction (sign) of change as one variable changes monotonically because of the interaction of another variable.

4.5.2 Statistical Power, Significance and Hypothesis Testing

Statistical significance is typically thought to relate to the likelihood that an observed result is or is not a reflection of the true underlying (unknown) distribution, population or system. However, the strict interpretation of the conditional probabilities involved in hypothesis testing reveals a contradiction: The likelihood of correctly rejecting the null hypothesis is conditioned on the null hypothesis being true (“given”).

This is examined further in Appendix D, but to return to introductory definitions, statistical power is the oft-neglected sibling of statistical significance. It relates to the setup of experiments of a statistical nature, and is intended to guarantee

the ability to discover an effect (a deviation from the null hypothesis) of a specified size at a specified statistical significance. For this, the choice of statistical power determines the minimum number of experiments to be conducted.

Hypothesis testing is the comparison of the estimated significance(s) of results to a previously established significance target. (Even without a prior target, hypothesis testing outputs a likelihood. According to some schools of thought, honest use of this likelihood to interpret the results of research requires a previously determined target-significance level. On the other hand, some of the problems with statistical significance are caused by over-reliance on the standard target level, which historically may have been a matter of mathematical convenience. This relates to another critique of statistical significance, which is that methods developed for the pencil-and-paper analysis of social or industrial data do not translate well to the computerized analysis of today's data sets.)

Leaving aside the concerns over the validity and applicability of hypothesis testing for the moment, it is also crucial to note that this comparison of target and calculated likelihoods requires different tests for different data types, different numbers of data sets, and the relationships sought among them.

Overall, then, hypothesis testing is a family of techniques that ask: Do different treatments (networks, hypotheses, algorithms) produce different outcomes? Since it is not feasible to study entire populations, regardless of whether they are populations of humans, viruses, artificial neural networks or audio recordings, sampling is a necessary component. Sampling introduces uncertainty, called sampling error. There is no blame or fault implied in the term “error” in this context—sampling error is part of nature.

But because there is always sampling error, we need a way to measure how much we ought to trust the results of studies of samples. The type-I error is the case of observing a difference (between performances or outcomes) when there is in fact no underlying difference. It is a false alarm, a false accusation, or a case of excessive credulity.

The likelihood of making a type-I error is the P value, which is a calculated, reported value, not a set target.

The corresponding target is the α threshold. One interpretation of scientific honesty requires the α threshold to be set in advance. Otherwise, any result could be declared statistically significant by retroactively picking an α value that would make it so.

If $\alpha = 0.05$, then $P < 0.05$ means “less than a one-in-twenty chance that the two samples emanate from a single underlying population” [132].

Related to significance is statistical power. “A study can produce results which are not statistically significant for two reasons. There may be no real differences, or alternatively, if present, the study may be insufficiently powerful to detect them. This latter possibility is known as a type-II error.” [143]. A type-II (or type-II) error, therefore, is the failure to observe a difference when the performances of networks, algorithms, or treatments truly differ.

The probability of making a type-II error is called β . Typically (if considered at all), $\beta \leq 20$, and we say that the study has 80% power. This means “if there is indeed a

difference of this given magnitude between groups, there is an 80% chance of correctly detecting it as statistically significant” [132, p. 1455] at the target significance level.

Statistical power ($1 - \beta$) is controllable by the experimenter. Online and other-software calculators exist, which allow one to find power given sample size, α and effect size, or find sample size given desired power, α and effect size.

When both significance and power targets are set⁷⁴, the effect size determined⁷⁵, and experiments are concluded, a statement along the lines of the following quote must be made for complete clarity as to the nature of the statistical finding:

There is at least an 80% likelihood that, had there been a 30% difference between groups, we would have found that difference with a value of P of less than 0.05. [132, p. 1456]

This topic is clearly challenging and confusing even to scientists and scholars. The misuse of statistical techniques in science, medicine (an applied science) and technology has been noted by many authors and researchers in the statistical, medical, computing and Machine Learning communities:

- Hastie, Tibshirani & Friedman⁷⁶ (2011) ‘The Elements of Statistical Learning’ [39]
- Siegfried (2010) *Science News* [132]
- Mlodinow⁷⁷ (2008) ‘The Drunkard’s Walk: How Randomness Rules Our Lives’ [145]

⁷⁴ $\alpha = 0.05$ and $1 - \beta = 80\%$ for the present study, after much deliberation, with the deciding factor coming from an excellent analogy in the Israeli Journal of Emergency Medicine, [144].

⁷⁵ The established effect sizes (small, medium, and large) are based on the extent to which the sought effect is greater than the null-hypothesis value, as a function of fractions of the variance.

⁷⁶ Stanford statistics faculty.

⁷⁷ Caltech physics professor and co-author of Stephen Hawking.

- Ziliak/McCloskey (2008) ‘The Cult of Statistical Significance’ [146]
- Ioannidis (2005) *PLoS Medicine* [133]
- Miller (2004) *The Journal of Systems and Software* [134]
- Zucchini (2000) *Journal of Mathematical Psychology* [135]
- Forster (2000) *Journal of Mathematical Psychology* [147]
- Salzberg (1997) *Data Mining and Knowledge Discovery* [33]
- Prechelt (1996) *Neural Networks* [136]
- Flexer (1996) *Cybernetics and Systems* [137]
- Holte (1993) *Machine Learning* [138]

According to Zucchini, when many competing models are evaluated in a study, random fluctuations in the data will increase the scores of some models more than others. The more models there are, the greater the chance that the winning algorithm or hypothesis wins by luck rather than by merit [135]. In a similar vein, Salzberg maintains that repeated searches of the same database with powerful algorithms will give rise to “discovering” nonexistent phenomena [33]. This is the Machine Learning version of seeing faces on the moon: With a powerful-enough pattern recognizer, spurious patterns will eventually appear salient. This is also partly because the statistical methods developed by Fisher, et al. were not designed for today’s powerful Machine Learning techniques, and their use in these settings can lead to misinterpretation. (This is a note of caution for the present research.)

In a study with similar goals and subjects, Holte reveals that by 1993, there were 75 published accuracy figures on 16 data sets in the UC Irvine repository. This, he claims, means that any more Irvine experiments will likely find statistical significance purely by statistical accident [138].

There are four problems with statistical significance and hypothesis testing:

- Public misperception,
- Misuse and abuse,
- Publication bias, and
- The meta-significance problem.

The public-misperception problem is that “statistically significant” sounds as if it means “significant,” which it does not necessarily mean. An example contrived for the purpose of clarification is the national averages of male and female test scores. If a well-designed (blocked, randomized, stratified as needed) and properly conducted study that compared the means of 100,000 male and 100,000 female students on a particular standardized exam found a statistically significant difference in mean scores of 2 points, this may be interpreted as implying that (at the significance level set prior to the collection of data) the observed difference in means *has a likelihood of not reflecting the underlying reality* that is lower than the threshold. It does not, however, mean that the underlying reality that was revealed has social or biological significance. The difference of two points may be well within the noise floor of natural variation in any test-taker’s health or concentration level on a given test day. Thus, the result is clearly *statistically* significant, and just as clearly irrelevant. (If the score difference were 100 points, the statistically significant difference would also be socially significant.) This distinction is, understandably, lost on most individuals who have not spent a great deal of time pursuing an understanding of it, and that includes many scientists, science editors of magazines and newspapers, lawyers, police and judges, doctors and patients, teachers, students, and indeed, most voters. The less expected, worrisome statistic, however, is

that 50 to 80% of leading-journal articles equate statistical significance with importance [137, 138].

Miller [134], Zucchini [135], and Salzberg [33] all tackle the multiplicity effect. One example [33] is of a study that compared 14 algorithms on 11 data sets (154 combinations). Each of these was compared to a default classifier using a two-tailed paired t test with $P < 0.05$. The problem is that this is not the correct test for this study. With this setup, there is *at least* a 99.96 % chance of incorrectly claiming statistical significance:

There are 154 chances for a result to be statistically significant. Thus, the expected number of significant results is $154 \times 0.05 = 7.7$. To calculate the proper α value for this study, let's first define α^* as the probability that if there is no true difference, we find at least one statistically significant difference.

Then, $1 - \alpha^*$ is the chance of getting the right conclusion per experiment

This, in turn, raised to the n th power is the *optimistic* chance of making at least one mistake. This is *optimistic* because the t test assumes independence, so the calculation above is valid under an assumption of n independent test sets.

The real alpha value is $\alpha = 1 - (1 - \alpha^*)^n = \sim 0.0003$. In other words, no p value higher than 0.0003 should be taken to indicate a statistically significant result. And if n distinct test sets were not used, the true α target is even lower.

The philosophical debates about the setting of such thresholds, the roots of how they were determined, and the use and reporting of these values are all lively topics

in the Statistics literature, and the question for the present researcher remains: What is the best way to go about conducting a rigorous study?

The answer offered here is to learn as much as possible about the statistical tools, controversies, viewpoints and alternatives, and to keep an open mind while working to the best of one's intellectual, mathematical and computational capacity.

For this reason, a roughly two-year portion of the present study was invested in the comparison of statistical model-selection criteria and performance-estimation methods.

The qualitative result of these studies is that all statistical tools and methods have strengths and weaknesses that are context-dependent. The competing tools include penalty measures like AIC and BIC (all of which have pros and cons with respect to one another), minimum description length (MDL), various forms of cross-validation, the bootstrap (resampling with replacement), the jackknife (the leave-one-out version of the bootstrap), the lasso, and many others. Various equivalences have been established in the literature: Holdout gives an unbiased estimate of generalization performance [147, p. 123]; AIC, LOO (leave-one-out cross-validation) and the bootstrap are asymptotically equivalent [41, 148], but LOO degrades as sample size grows [149], and overfits in model selection [42]; k -fold cross-validation is superior to both holdout and LOO [30]; 10-fold cross-validation is better than any bootstrap [43, supported by 44], but stratified k -fold is the best, and k can be made smaller as n increases [150].

Hence, the choice of 5-fold stratified cross-validation for the fully connected NNs in the present study when determining the number of hidden-layer elements.

The preceding section's findings were the result of awareness about the problems with hypothesis testing and statistical significance, to which we return.

The misuse problem discussed above is on the generating side of statistical-significance data and claims: Scientists, applied scientists and engineers often make unsupported assumptions or use the wrong statistical tests to conduct on their data, and thus mistakenly pronounce finding or not finding statistically significant results.

Publication bias is the unfortunate but natural situation that arises when many scientific studies on the same topic (all assumed statistically sound) are submitted for publication to a small number of prestigious journals during a short period of time. If four out of ten studies discovered statistically significant results at a given target rate (that is common to all studies, for the sake of argument), and six did not, it is likely that some or all of the four will be published, and some or all of the six will be rejected.

Gould explains the phenomenon thus:

Only the most miniscule proportion of scientific studies ever get reported in the press, and these decisions often bear little correlation with the importance of such studies for professionals. Better relationships can be found between the decision to report and the degree to which a conclusion disturbs conventional notions (often misconceptions) about the nature of things. (Gould, [151, p. 208])

A more telling situation is if more than twenty such studies were carried out, and only one resulted in findings at a statistically significant level. When that lone study

gets published, the act of publishing that study is itself not statistically significant at the typical 5% level⁷⁸.

Negative results are not appealing to the human psyche. But they are just as important. Knowing that statistical significance does not correlate fully (if at all) with importance, it is imperative for the healthy functioning of the scientific endeavor for negative results to be published.

The meta-significance problem is that there are challenges within the statistical literature to the very idea that statistical significance ought to be pursued at all. This dissertation does not take a stand (as it is not in the field of Statistics), and indeed pursues statistical significance, but also recognizes the argument against the practice [133, 113, 152]. The crux of the matter is that a very small change, potentially a statistically not-significant change, can move the results of an experiment from turning out significant at a target level, or not. This is partly because certain target significance levels are typically chosen, such as 0.05. The result may well be significant at the 0.06 level. The question is: Why should all Science revolve around 95% likelihood and reject 94% likelihood? And then again, at the 0.06 level, the question may be: Why not 0.07?

Other criticism of hypothesis testing includes the accusation that scientists mindlessly spend their time repeating experiments to achieve statistical significance for results they might know to be false or useless [113], and on the other hand, reject useful findings because the experimental design or natural sampling error did not lead to a

⁷⁸ See <http://xkcd.com/882> for a humorous demonstration of this.

significant result. At the same time, significance by accident is just as common [135, 138].

There is an element of arbitrariness to hypothesis testing and significance that is of concern. However, the solution of such matters ought to be left to statisticians and philosophers of science. The aim here is to present both enlightened opinions. In either case, hypothesis testing (in the present study, Dunnett and Welch's ANOVA, the Bonferroni's adjustment, Tukey's test, Hsu's comparison, etc.) based on prior consideration of statistical power and significance is a better tool than not testing at all.

4.5.3 Control Design

The network structure that serves as control has arbitrarily selected inputs that are neither the result of a modeling effort nor intuition about the problem domain, but are intentionally chosen to be meaningless. (This is the "lousy-control" idea: Design an intentionally lousy network. All meaningful designs must outperform this control to be taken seriously.)

The performance of a network prestructured in this way serves as a baseline for *accidental* generalization performance, and any performance measurement of intentionally prestructured (or even fully connected) networks must overcome this performance to a significant degree in order to be considered successful.

This structure was determined by repeated use of a random number generator, in the same way as tossing a die or coin multiple times. This method was chosen over "just making up random connections" in order to avoid any subconscious bias on the part of the designer.

The first “toss” determined how many hidden elements to have. The range was 0 to 50, and the outcome was 4.

The next four tosses determined how many inputs should connect to each hidden element. With a range of 0 to 17, the outcomes were 3, 2, 13, and 7.

The next long series of tosses determined which inputs to assign to the above connections. The random sequence was 7, 3, 7, 8, 3, 10, 16, 0, 12, 16, 13, 8, 16, 14, 1, 11, 12, 15, 7, 3, 16, 10, 8, 7, 4, 16, 15, 4, 3, 4, 11, 4, 9, 4, 3, 12, 15, 11, 9, 12, 12, 5, 7, 9, 12, 5, which resulted in the following components: CGH, CK, ACDGHJKMNPQRS, and CDEJMNR.

The resulting “lousy model” is IV:CGHZ:CKZ:ACDGHJKMNPQRSZ:CDEJMNRZ. This model was trained and tested just like any other model at the holdout-test stage.

4.5.4 Assumptions, Justified and Unjustified

An important issue in the use of statistical models or methods is the justification of the assumed underlying distribution. Perhaps because almost all academic examples are based on Gaussian (normal) or uniform distributions, researchers sometimes assume without justification that one of these two probability distributions must apply to their data. Most generating processes, however, are unknown, and it is only safe to assume a Gaussian distribution when composite experiments are performed, leading to accumulation of “large” numbers of data.

In the present research, no claims have been made as to the underlying distribution for RA models, neural-net performance, or the occurrence of clave-related patterns in musical practice.

4.6 Experimental Process

4.6.1 The General Experimental Process

The experimental design is determined by several decisions. Among these is the conclusion of the statistical research that indicates the use of stratified 5-fold cross-validation, 393⁷⁹ random-number seeds (listed in Appendix O) for statistical power under the present conditions, a preference for AIC in loopless searches and BIC in all-model searches, and the use of Dunnett and Welch’s ANOVA for the final comparison of fully connected networks, prestructured networks and RA percent-correct on the holdout set. In addition, the experimental design is informed by the lengthy period of exploratory experimentation, which led to the choice of certain training-and-test regimes, namely training with STRONG⁸⁰, AVE, and all (specifically, design-all) vectors, and testing on the appropriate (meaningful) combination of AVE or all vectors.

⁷⁹ This number is due to Cohen [70], and refers to the number of instantiations required by statistical theory for the statistical power to find even a small effect at a significance target of 0.05 (α) in a two-way comparison. A three-way comparison would require fewer repetitions (322), as would the statistical power necessary to find a larger effect [61, p. 158]. This is disputed by Mathematics professor Lenth in a paper not yet published, called “Seven habits of highly effective sample-size determination” (submitted 2000).

⁸⁰ “STRONG” includes those category-1 and category-2 vectors marked “strong” and “very_strong” just as “WEAK” includes all category-1 and category-2 vectors marked “weak,” “very_weak,” “WBD,” and “VWBD.” For categories 0 and 3, the annotations are divided up into STRONG, WEAK and AVE such that, for instance, “fully_neutral” and “GOOD_zero” go into STRONG, “neutral” goes into AVE, and “better_1” and “better_2” (two types of weak neutrals) go into WEAK.

Furthermore, the factorial aspect of the design requires the preceding steps to be repeated for the two separate holdout conditions and for the two one-output-at-a-time output encodings (one for forward and one for reverse clave direction). In the interest of ever completing the research, the random holdout and the forward output encoding were selected as the first combination to investigate. Only if satisfactory results were not achieved would the other three combinations (or yet further different encodings) have been investigated.

Stratification for a data set may proceed as follows. The training set of “design-all” vectors from the random-holdout set consists of 8631 vectors. In the forward OOT encoding, these are either forward (output 1) or not (output 0 for reverse, neutral or incoherent).

However, in addition to the output categories, the data are drawn from an organized collection where vectors with strong, average, or weak membership in their clave-direction categories are in their respective groups (for maximum control over and knowledge of the data). Stratification, then, must concern the mix of these membership degrees as well as the output categories. The reason is clear when we imagine what would happen if membership degrees are not stratified in 5-fold cross-validation: Some training or test sets would get their examples exclusively or mostly from a group of strong examples while others would get their content from weaker sets. As a result, each fold would not represent how a network randomly presented *all* this data would do when generalizing to an unseen new set. Since this is the purpose of cross-validation,

the splits must be composed so as to allow the cross-validation to succeed in predicting future generalization performance.

One way to achieve this stratification is to split the design data into an output-1 group and an output-2 group, and then shuffle each group within itself before taking off the top fifth for the first fold's test, the second fifth for the second fold's test, and so on.

The preceding steps are the core of the experimental procedure, but further steps could be possible and may become necessary, depending on the outcome.

For each training-set grouping, a parallel series of RA searches, and to the extent possible (based on response or lack thereof), parallel human floor and ceiling trials are conducted⁸¹.

When a set of networks are trained on, say, all design vectors, a 5-fold cross-validation split is used to determine the optimal network size (for the default random-number seed). Once the size of the fully connected network is determined, 273/44/17 other seeds are used for initialization, and the results are recorded for thresholding and statistical comparison.

At the same time, the same training set of all design vectors undergoes RA searches to determine the best prestructuring model (through the heuristic search strategies introduced in Section 1.16 and exemplified in Section 4.6.5). The structure of each prestructured network is determined by the final choice of RA model, so no

⁸¹ Response indeed has been too slow to keep up with the other streams of the research. Several participants dropped out (one from the expert group) and those who took part have taken on the order of six months to complete a training-and-test run, so only the STRONG vectors have been through the human benchmark.

further tuning—no cross-validation design—is necessary. Such a network structure is then exercised with 274/45/18 seeds (the same set used for the fully connected net).

The Generalizing Ratio (GR, Section 2.2.11) is calculated for each instantiation, and the means of different hypotheses' GRs are compared in a Dunnett and a Welch ANOVA for the final conclusion.

4.6.2 The Processes of Data Acquisition and Preparation

Under traditional practice, the patterns that are consciously brought up (played, designed, written, executed) by the keepers of the tradition tend to fall into relatively clear examples of the traditional concept of clave direction according to *partido-alto*-based *samba carioca*. It would be very rare that the incoherent and the weakest examples of this, as found in the present design data set, would arise with any appreciable frequency in musical practice, but as the trainer of machines (and the one who is expanding the concept of *partido-alto* direction out to all possible rhythmic situations), one has to consider what would happen when these unlikely cases arise⁸², based on a small and inconsistent sample of responses from master teachers⁸³ to a sampling of these patterns, presented to them in such a way as to test for the consistency and reliability of their responses. Since such reliability and consistency was found to be lacking (which is natural, considering the most difficult patterns are those that are least

⁸² As Baroni explains in *The Concept of Musical Grammar* [153], “the aim is to arrive at a set of rules that not only possesses coherence and concision, but also the necessary comprehensiveness, that is, one able to account for all phrases composed — *or which one might want to compose* (emphasis added) — in the style of a particular repertory” [153, p. 186] which is “the classic prerequisite of Chomskian grammar” [153, p. 201].

⁸³ Teachers who are master-players, and possibly but not necessarily master-teachers.

likely to occur in the idiom), the researcher is tasked with defining and creating a *consistent* framework (*based* on the master teachers' responses and their teachings that the researcher has been exposed to) that is more general than any existing tradition or literature. Such a framework is needed (possibly even multiple such frameworks are needed) because machines are to be taught to discern this cultural concept, and they cannot benefit from the cultural insights and experiences of even the least exposed human novice.

Under the worst case, it would still be acceptable for humans to throw up their hands and say "I don't know." And while undecided outputs are acceptable in machine learning, doing so with hundreds of patterns would reduce the design data to statistically insignificant numbers.

To make the data representative of a consistent and culturally informed interpretation of the concept at hand, instances where the most important criteria (derived from studies as well as through indirect testing or direct questioning of teachers) disagree with one another have been repeatedly cross-checked until a consistent framework, however unique it may be to the researcher, was attained.

A priority order was generated for nine criteria identified through studying and analyzing this musical tradition from 2001 through 2010. Four of those criteria are ranked as paramount.

When all possible patterns are considered, it is expected that many will not conform to the notion of clave direction. Among those, it is not at all unusual to find a

pattern for which one or more of the paramount criteria give one answer with high certainty, and others give the opposite answer with near-equally high certainty. An example of this was the topic of an SMS exchange (text messages via cell phone) on Monday, September 20, 2010, between the present author (A) and one of the master teachers (T⁸⁴), a significant informer of the present research. The exchange was initiated by the latter. The messages follow (verbatim, in fixed-width font to facilitate interpretation of this version of TUBS notation):

T: "here's one: 2-3 or 3-2? e+a2 + e 4 a"
(11:57)

A: "Good one! I think both Monobloco and Spiro would call it 3-2, but for me, it's "incoherent" (neither, not neutral)." (13:01)

T: "as a variation of: e+a2 + 3 + 4 a, i agree 3-2 ... something about that "+" of 2 ... but if we're talking "tension and release" it seems more 2-3?!" (13:06)

A: "Yes, I can see that. I do think there are multiple aspects to the idea, so when you have unusual (non-traditional) patterns, you gotta identify the context, i.e., instruments, tempo, role ..." (13:12)

T: "right ... and is it a variation, or a 'part'?" (13:14)

The final response brings up an important concern which has been addressed in the data-classification process by defining separate teacher models: Among other

⁸⁴ Derek Reith, profiled in DRUMHEAD Magazine, No. 29, September–October 2011, New York, NY: PPV Media, LLC, p. 43.

aspects, the teacher models address whether to interpret a given pattern (a snapshot, if you will) as repeated or not repeated. The strict teacher model assumes that any pattern presented to it may be an ostinato, and hence classifies it under the strictest interpretation that the author has encountered, resulting in a higher percentage of category-0 (incoherent) patterns than the other teacher models. The firm teacher model also makes the same assumption (but relaxes certain other criteria). The lenient teacher model does not assume that a pattern is an ostinato—it may be played only once, as an introductory figure, an ending figure, or part of an improvisation. As Spiro suggested, its clave direction may depend on what preceded and what followed it in a musical setting. The inspiration for this model came from comments recorded during Spiro’s data-acquisition session when he said that several of the figures played behind the curtain would be acceptable in the given clave direction, but only if they were “entrances” or one-off figures.

4.6.3 Data Acquisition: Classification

“[I]f you are not following the spatial succession of marks on a score, but rather following the succession of events as they occur in real time, the ‘same thing’ when heard in a different context can sound very different.” (Bamberger, [154])

There are multiple steps to the data-acquisition process. Since it was known at the start of the research process that 16-bit binary strings would represent the inputs (rhythm patterns), the first task was to generate all such strings (vectors). Then, the vectors were shuffled using a high-quality random-number generator.

After this, in September 2006 began the process of classifying the vectors into the forward, reverse and neutral clave categories. The first important insight was that some vectors were similar to traditional patterns with known (accepted) clave directions from cultures other than the samba of Rio de Janeiro, but did not fit all the characteristics of Rio-style samba patterns traditionally labeled as being in the same direction. These other styles included Cuban rumba and Brazilian music from various northeastern states. The author's concept of clave direction had to be broadened, tested, improved, and more often, narrowed down and made more specific until the iterative process of going back and comparing patterns that had a similar characteristic with a new one under consideration eventually led to the types of questions for which the researcher's knowledge and authority were insufficient. These questions (patterns) were taken to several of the master teachers (cultural experts) and mid-level teachers in a series of double-blind experiments (except in one case where the expert was able to simply look at and mark the vectors).

For the double-blinded experiments, the researcher was on one side of a screen where he could not see the subject (expert) and where the subject could not see the researcher. On the other side of the screen were the subject, two reliable drummers (*surdo* and *caixa*), and two note-takers. The drummers played the Mocidade samba school's standard samba, which indicates clave direction very clearly and definitely. The researcher on the other side of the screen played each problem pattern on a *tamborim* (a small but loud frame drum), repeating a given pattern until asked to stop. The patterns were arranged so as to include 180° shifts

of almost every pattern (as one type of test of intra-rater reliability) and some “obvious” or “easy” patterns (as controls).

The note-takers recorded not only each expert’s statements as to clave direction, but also all comments and gestures, in order to aid in interpreting results in cases where experts disagreed (which for these “difficult” unusual patterns happened many times). For example, the pattern 0100 | 1010 | 0100 | 0101 got the “OK” from one expert, “no” from another (as to whether it conforms to the forward direction being played by the accompanists), and a non-committal descriptive response from another. (The fourth expert got a difference set of patterns created based on what was learned from the first three, due to the time gap between sessions.)

Another example, 1000 | 1100 | 0100 | 1100 is representative of the type of result that—after further consultation with some of the master teachers—led to the idea of different models of teacher strictness. This pattern is a clear example of the reverse clave direction (and three of the master teachers said so), but is so far from typical practice and typical African or African-derived patterns (and so close to neutral timekeeping) that the most esteemed master teacher said that it was “OK” over 3-2 (forward) samba.

Several examples of patterns on which all experts *and* the author’s clave-direction theory as it stood at the time of the experiments include the following:

- 0101 | 1010 | 1011 | 1101 (very strong forward)
- 1001 | 0001 | 0100 | 1010 (strong reverse)
- 1011 | 0110 | 0001 | 1011 (weak-but-definite forward)
- 1001 | 0010 | 1001 | 0100 (strong forward)
- 1100 | 1010 | 1010 | 1010 (average forward)
- 1111 | 1010 | 1010 | 1111 (strong forward)
- 0101 | 1010 | 1010 | 1111 (very strong forward)
- 0100 | 0101 | 0100 | 1010 (very-weak-but-definite reverse)
- 0110 | 0010 | 1001 | 0100 (average forward)
- 1101 | 0100 | 0010 | 1011 (weak forward), and
- 0101 | 0000 | 0011 | 1010 (very weak forward).

This last example is among the patterns that informed a critical aspect of the currently proposed theory of *partido-alto* clave direction: the relative weight given to classification criteria. The first part of this pattern, 0101, is a very strong indication of forward direction when in the first position. However, the last part, 1010, is a rather strong indicator of reverse direction when in that position. Because the remaining sections of the pattern confer almost no information as to directionality, the decision (in which all experts and the author agree) is that the schema 0101 is the most powerful schema in samba, and thus overrides the effect of the less significant 1010 schema.

As for the mid-level teachers' experiments, two of the sets of their results were disqualified from consideration due to mismatch with the controls. For example, the pattern 0000 | 0111 | 1101 | 1010, which is a strong reverse (2-3), was called "kinda 3-2" by one subject (mid-level teacher 2), who, when presented with 1001 | 0101 | 0101 | 1010 (very strong reverse), said "figuring it out ... kind of in clave" without stating what direction. Although slightly tricky (due to the 1001 start, and only slightly), this pattern is a very strong example of conforming to 2-3

partido-alto, not an average or weak example as suggested by the expression “kind of in clave.” Mid-level teacher 1 also got hung up on one of those two controls, and gave irrelevant answer to several other patterns.

After the first set of experiments with the first three master teachers, the results were compiled. Responses were first considered to test for consistency among the experts. Those patterns for which all masters agreed were used the main basis for improving the author’s hypothesis of clave-direction in *samba carioca*. Patterns on which the experts disagreed were used in further investigation with the remaining master teacher, and helped solidify aspects of the theory like the different teacher models and the five membership degrees in a category (which together code *some* of the cultural dependency of clave into the data-acquisition process), and also for the related but separate goal of developing a cross-cultural theory of clave. In such cases of disagreement, the comments notated by the note-takers were instrumental in finding out when the experts were thinking of other cultural contexts, which they all indicated at one point or another, referring to *marchinhas*, *maracatú*, and various Afro-Cuban contexts.

After this, the next two or three passes through what was then only about 6000 vectors were used to hone the researcher’s internalization of the ideas hidden in the experts’ answers. Over the course of about two years, the present researcher re-classified, reviewed, and cross-checked the data set, frequently using computer-sequenced samba ostinati generated in a program optimized for non-Pop/Jazz/Classical percussion: Henry’s Percussion Studio (moosware.net/PercussionStudio),

which implements a form of TUBS [155] for its sequencer, but goes beyond TUBS notation to allow expressive timing through the sensitive placement of note triggers in non-quantized positions.

At the end of two years of multiple-cross-checking, the list of criteria previously mentioned were obtained, along with precise notions of the meaning of STRONG, AVERAGE, and WEAK in each teacher model (strict, firm, lenient).

4.6.3.1 Criteria for helping determine the *Partido-Alto* Clave Direction of Challenging or Vague Patterns

In the discussion below, the term “up” indicates high local offbeatness, and “down” indicates low local offbeatness. The criteria are listed below.

1. The *Partido-Alto* Form
2. Isolated Missing Downbeat (on 1 or 3; IMD)
3. Gross Relativeness (the main concept of relativeness in Appendix A)
4. The third *surdo* from Mocidade (a partial form of *partido-alto*)
5. Fine Relativeness (involving the relative strengths of nibble-length schemata, as shown below, and the “algebraic” cancellation of onsets across schema bridges, first to third, and second to fourth)
6. Template-Matching against the *teleco-teco* (assisted by the shortcuts “zipper” for 3-2 and “bookends” for 2-3)
7. The “African” Clave Sense (antecedent/consequent relationship)
8. The Partial (“Hanging”) version of IMD
9. Direct Template-Matching to “bossa clave” (only when no other criteria, including “how it feels” lead to an answer, such as in
1111 | 1111 | 1111 | 1100)

Criterion 1 has top priority. Criterion 2 is the second most powerful.

Criteria 3 and 4 are of equal weight, followed by Criterion 5. Criteria 6, 7, and 8 have equal weight, and Criterion 9 is the last resort when none of the others lead to an answer. (This can happen either due to those criteria not being relevant, or due

to conflicts among them, although the latter typically leads to the “incoherent” category.) Some examples follow, in order to illuminate what is meant by these criteria.

0100|0100|0101|0000 is a very weak reverse pattern under the firm-teacher model: Criterion 1 does not apply. Criterion 2 gives equal likelihood (and less weight, due to H|JK being 0|01, as opposed to 1|01, which would confer full weight) to forward and reverse. Criterion 3 indicates reverse because the gross schemata are up|up|up|none, meaning offbeat-heavy for the first three, and empty for the fourth. Criterion 4 suggests reverse because of good match with 2-3 *surdo* following the third beat. The other criteria do not contribute.

0111|1101|0110|1110 is a strong incoherent under the firm-teacher model: Criterion 1 doesn’t apply. Criterion 2 gives both a full indication of reverse (1|01 surrounding beat 3, or J) and a partial indication of forward (0|01 around beat 1). Criterion 3 gives a hint of reverse due to the schemata being up|up|indeterminate|down, where the first two cancel each other out. Criterion 4 gives a strong indication of reverse because of perfect match with 2-3 *surdo*. The only other criterion that applies is Criterion 6, where the “zipper” idea (inspired by textbook images of RNA transcription to describe certain ways in which cleave direction is manifested) gives a slight sense of forward in the way the & of 3 (M) and beat 4 (P) match *teleco-teco*.

Likewise, 1110|1111|1111|0101 is a strong incoherent due to votes of forward, nothing, reverse, forward, forward, reverse, nothing and nothing by the criteria in respective order.

As obvious in these examples, this set of criteria alone is not sufficient. The prioritization of the criteria is a form of codification of human intuition and social/musical conditioning regarding these patterns, and the method of combining the suggestions made by these criteria is still up to the observer/listener/performer.

Furthermore, these criteria are only for use with very challenging, unusual patterns. Those vectors for which the clave direction (or lack thereof) is obvious do not require this type of intellectualization and analysis. The need for this system of prioritized criteria was developed because patterns can be appear arranged in such a way that phrases go across “tactus boundaries” of the schemata, and the overall offbeatness or onbeatness of each phrase (relative to one another) is more relevant than the first-, second-, third-, and fourth-position 4-bit schemata. All the criteria except the third and fourth incorporate some musical and less mechanistic aspect of the notion of clave direction.

Should the first two criteria not be thoroughly violated, and the third and fourth are pulling in opposite directions, the relativeness of schemata can be used as tie-breaker, but only in the firm-teacher model.

4.6.3.2 The Case of the “Average” Membership Degree:

The term “average” has been used in this research effort to mean three different clave conditions with similar results:

1. Clear evidence for the given clave direction, but there's only a little bit of it (with nothing else pulling otherwise);
2. Strong evidence for the given clave direction, but there is also a weaker pull in the opposite direction, so the resultant is average in its strength;
3. Highly representative in-between cases, which display “garden-variety relativeness” where clave direction can be revealed by an onset-by-onset comparison of the inner and outer regions.

Average is also when schemata are pulling in opposite directions, but the pull in one direction is definitely not as clear as in the other, such as:

1010|1110|0101| 1111 with down|down|up|up schemata. This is not an incoherent pattern (in the firm context) because the first and third schemata, which together reinforce the reverse characteristic, are rather strong indicators, whereas the second and fourth schemata, though both point toward the forward direction, are weaker indicators.

4.6.3.3 The Three Teacher Models:

The differences between the firm and strict contexts come down to a few details. In the firm context, the *partido-alto* form is not the absolute ruler, whereas in

the strict context, any violation of the P-A form leads to a class 0 (incoherent) output. The lenient context is mainly “receptive,” not necessarily “generative,” meaning that a pattern that does not serve to *establish* a clave direction, but also does not preclude either direction can be accepted because it can accompany other patterns that are in clave without harm.

Thus we find out that in the firm context, violating Criterion 2 (when 01 starts beat 1 or 3) does not lead to an automatic incoherent classification when it happens on both sides, or when it happens with a weaker schema than 0101 (which are 0100, 0110 and 0111). (The schema 0100 is only included solely because of Rio supergroup Monobloco’s common repetitive use of such patterns.) Criteria 1 and 8 can be of help with this, but if those criteria are not conclusive, the pattern needs to be considered very carefully.

Any similar but weaker schema, like 0001, is not considered to be involved in Criterion 2, though, as always, the final result depends on the rest of the pattern, as in 0001 | 1010 | 0101 | 0101, which is a strong incoherent in all three teacher models.

An important exception to all criteria is that, under firm and lenient, fully neutral patterns are exempt from all other consideration. Hence, even 0101 | 1010 | 0101 | 1010 is neutral under firm and lenient (but never under STRICT). Also, in the lenient case, Criterion 3 may have greater weight than Criterion 2. (This is one of those musical reasons arrived at through “internally

listening” a very large number of these rhythms.) It is situations like this that make neural networks good candidates for clave-direction recognition: No matter how carefully we extract precise rules, exceptions exist because “that’s just how it feels.”

The following entries show how the same pattern can be classified differently under the firm and lenient contexts, with the firm result followed by the lenient.

0100 0100 0101 0000 strong neutral; very weak reverse

0100 0100 0101 1000 strong neutral; very weak but definite rev.

0101 0101 0101 0011 strong neutral; very weak reverse

When Criteria 1 and 2 are in clear conflict, which can happen in two main ways, the two remaining schemata (beats 2 and 4) can have two strong indicators, each agreeing with one of the two conflicting schemata of Criteria 1 and 2, or they may fail to provide any help. (In either case, they fail to provide *consistent* help.)

A *weak* incoherent is when the *only* reason for being incoherent is an alternating sequence of schemata (such as up | down | up | down, like one of the riffs in Glenn Miller’s *In The Mood*). An average incoherent could be a result of up | up | up | up or down | down | down | down, as long as there is no musical reason to reject the pattern more strongly.

We close this section with an example of the counterintuitive nature of P-A clave direction: 0010|1000|1010|1111, which comes out to be forward (although it starts out sounding exactly the same as the 2-3 *son* clave) by criteria 1, 3, and 5.

For further detailed analyses of the classification and membership degrees of all possible patterns under different teacher models, contact the author.

4.6.3.4 Some Commonly Recognized Schemata:

Recall that “up” implies high offbeatness. An up schema in first position (beat 1) or in fourth position (ending just before beat 1) contributes a strong forward sense. Likewise, “down” indicates low offbeatness, and contributes a strong forward sense in the second and third positions, and vice versa.

Very strong indicators:

- 0100 (up, but only indicates *very* strong in first or third position, less strong in second or fourth position)
- 0101 (up)
- 1010 (down)

Strong indicators (schemata):

- 0010 (down)
- 1101 (up)

Medium-strength indicators (schemata):

- 1000 (down)
- 1001 (up, or neutral *only* when repeated)
- 0111 (up when it is in first or third position; weaker or down when in second or fourth position, all relative to what else is around)

Weak indicators:

- 0001 (generally up, but highly dependent on what else is around)
- 1110 (down)

Very weak indicators:

- 1111 (up, but occasionally not!)
- 1011 (down, but occasionally not!)
- 1100 (up, but occasionally not!)

Non-indicators:

- 0000
- 0011
- 0110

The Case of the Leading 0110:

With 0110, Criterion 2 is of lower priority than relativeness (Criteria 3 and 5) *unless* the previous (leading/hanging) 16th note is expressed. An example of the former case would be 0110|1111|1001|1110, which is a very weak reverse in the firm-teacher model, but incoherent in the strict-teacher model. An example of the latter case is 0110|1100|0100|0011, which is incoherent under all teacher models

4.6.4 Training-and-Test Regimes

The term ‘regime’ will be used to describe the selection of the factors that go into the design of experiments in the main thrust of this research. This term seems more appropriate than either ‘scheme’ (which has a connotation of dishonesty) or ‘strategy’ (which has a connotation of aggression).

Each training-and-test regime consists of five choices and five actions.

The choices are 1) output encoding, 2) training categories, 3) test categories, 4) training membership degrees, and 5) test membership degrees. The actions are:

- Data preparation based on the above choices and information from Lendaris, Haykin [26], Leen [29], Hastie, Tibshirani & Friedman [39], Mitchell [72], Cohen [156], NeuralWare;
- Running the trials,
- Parallel RA modeling,
- Thresholding the results, and
- Making comparisons with a control.

A large number of trials using different, randomly acquired random-number seeds (initialization seeds) must be run for each regime. This allows networks to be initialized to different parts of the search space, and sets the stage for seeking statistically significant results as well as high statistical power, and allowing the calculation of confidence intervals.

There are established guidelines for (fully connected) neural-net design governing such design choices as the range of acceptable numbers of hidden-layer elements, the ranges of momentum and learning rate, learning schedules (how parameters are varied during the learning process), and whether and when early-stopping is used.

In order to evaluate the effectiveness of information-theoretic prestructuring, there must be a statistically sound comparison with the standard method (fully connected MLPs) and the informed bracketing of expectations with upper and lower limits based on controls.

For the comparison to be valid, fully connected networks must be the best that can be designed within reason (the design process for neural nets is open-ended), following standard optimization procedures, but foregoing optional or cutting-edge

procedures specifically put forward in competition with the standard method (such as evolving neural nets or Bayesian design methods). This sets up a comparison that is fair on both sides, and is practical. (The comparison of all known neural-net optimization strategies would take an infinite amount of time, as would the investigation of all possible standard heuristics.)

Thus, the momentum and momentum schedule, learning rate and learning-rate schedule standard for MLPs on NeuralWare's NeuralWorks™ are used without variation. Similarly, the standard processing-unit type and learning algorithm are employed without venturing into Cascade Correlation, Logicon™, Delta-Bar-Delta or other refined MLP variations.

Human interventions like early stopping, pruning, and the bumping or jogging of weights are also avoided because the purpose of information-theoretic prestructuring is to make such interruptive measures unnecessary.

Fully connected networks (here, multilayer perceptrons) are designed following a combination of industry-standard heuristics. We start with a 5-hidden network and a 25-hidden network, and exercise each one five times with the five data splits of 5-fold cross-validation, as described in Section 1.8.5.

4741 vectors in the STRONG group, encoded for forward direction (one output at a time, meaning forward is 1, not forward is 0, without distinction as to what category if not forward) were stratified and divided into five training/test folds.

The 5-hidden network's numbers of correct were 911, 912, 922, 928 and 929 (out of 948 test vectors).

The 25-hidden network's numbers were 905, 911, 924, 929 and 930.

This was followed by a 15-hidden network (931, 928, 924, 915 and 906) and a 1-hidden network (930, 928, 913, 887 and 892). These four networks have cross-validation estimates of their generalization performance 97%, 97%, 97% and 96%, respectively.

The STRONG recognition problem appears too easy to even need to be tackled through prestructuring. (Nonetheless, the binary search for the optimal network size was continued, and resulted in the choice of a hidden layer of three elements, with 97% estimated generalization and the smallest number of elements at that level.)

After finding and removing a small number of duplicate vectors, the complete research data set had 8631 design vectors and 2153 holdout.

Again, the first step was to construct 5-hidden and 25-hidden networks. The averages of the five splits for both were 92% correct, though the outputs of the last split of the 25-hidden network featured oddly clustered values. (The 1562 correct out of 1727 test cases all grouped themselves into three or four output values.)

Since the two networks showed no difference in estimated generalization, the next step was to build a 1-hidden network (to move in the direction of greater simplicity). The 1-hidden network had just under 91% correct, but its values were also clustered, and this time for all splits. This network is likely not a useful design for later stages—it appears to be doing vector quantization, which is a useful feature, but not what is sought at this stage.

Halfway between the 5-hidden and the 1-hidden (binary search) is the 3-hidden network. Its performance estimate was also 92%, and its values showed healthy

variation. The performance of the 2-hidden network rounded off the binary search, which was 90%. Once again, the preferred size for fully connected networks is 3 hidden-layer elements.

This ends the all-category one-output-at-a-time (forward) fully connected design problem. What remains is the statistical evaluation of the fully connected network through the use of the predetermined number of re-initializations (followed by training and testing) with a set of randomly determined seeds for the final analysis.

4.6.5 RA Search, Fitting and Selection

As described in Section 1.16, there are two search types and two search-criterion groupings, resulting in four modeling approaches available to each training-and-test regime. These are enumerated in Table 7 according to their relationships.

The all-vector searches for the random-holdout, one-output-at-a-time, firm-teacher forward case are given as an example in this section. The different types of searches and the steps they consist of are abbreviated here as in “MBR—a1” (step one of search MBR—a from Table 7). The reader may refer back to Section 1.16 for the steps that make up each search type, and to Section 1.8.6 for the explanation of why some steps differ in their selection criteria.

Table 7: Combinations of RA Search Strategies and Criteria

Mixed-model bore–rotate statistical search with AIC, BIC and feature selection (MBR–a)	Mixed-model bore–rotate information-theoretic search with Information and feature selection (MBR–i)
All-model bore–expand statistical search with BIC without feature selection (ABE–b)	All-model bore–expand information-theoretic search with Information without feature selection (ABE–i)

4.6.5.1 Design-All, One Output at a Time/FWD, MBR–a with Random Holdout

In MBR–a1 (Figure 26), a narrow search in loopless-model space is conducted to determine how many variables can be justifiably included in the next step, which is the selection of features (input variables) to be included and excluded.

The resulting AIC model was IV:ABCFGJKMRZ (level 9) with a Δdf of 511 (acceptable in comparison to other possible models) and $\alpha = 0$ (no chance of a type-I error).

This informed the next step, MBR–a2. The level setting was 10 rather than 9, for some headroom that costs hardly any excess computational load. The width was set to 20 (very high for loopless models), as seen in Figure 27.

Composition Method:	<input checked="" type="radio"/> Standard <input type="radio"/> Back Projection (Fourier)
Reference Model:	<input type="radio"/> Default <input type="radio"/> Top <input checked="" type="radio"/> Bottom <input type="radio"/> Starting Model

Search Settings	
Models to Consider:	<input type="radio"/> All <input checked="" type="radio"/> Loopless <input type="radio"/> Disjoint <input type="radio"/> Chain
Search Direction:	<input type="radio"/> Default <input checked="" type="radio"/> Up <input type="radio"/> Down
During Search, Sort By:	<input type="radio"/> Information <input type="radio"/> Alpha <input type="radio"/> % Correct <input type="radio"/> BIC <input checked="" type="radio"/> AIC
When Searching, Prefer:	<input checked="" type="radio"/> Larger Values <input type="radio"/> Smaller Values
Search Width: (models to keep at each level)	<input type="text" value="2"/>
Search Levels:	<input type="text" value="17"/> (leave blank to use settings from data file)

Report Settings	
In Report, Sort By:	<input type="radio"/> Information <input type="radio"/> Alpha <input type="radio"/> dDF <input type="radio"/> Level <input type="radio"/> % Correct <input type="radio"/> BIC <input checked="" type="radio"/> AIC
In Report, Sort:	<input checked="" type="radio"/> Descending <input type="radio"/> Ascending
Include in Report:	<input checked="" type="checkbox"/> H <input type="checkbox"/> dLR <input checked="" type="checkbox"/> Alpha <input type="checkbox"/> % dH(DV) <input checked="" type="checkbox"/> dAIC <input checked="" type="checkbox"/> dBIC <input type="checkbox"/> BP-based Transmission <input checked="" type="checkbox"/> Incremental Alpha <input checked="" type="checkbox"/> % Correct <input type="checkbox"/> % Coverage of Data <input type="checkbox"/> % Missing in Test

Figure 26: The settings for a narrow *Occam3* RA search (MBR–a1) of mathematical models with statistical model evaluation and selection, to determine the number of variables (which correspond to “levels” in a loopless search) for the subsequent step, along with provisions (in the data set and the search settings) for RA generalization results to be used in the ultimate comparison that is the goal of the present research.

Composition Method:	<input checked="" type="radio"/> Standard <input type="radio"/> Back Projection (Fourier)
Reference Model:	<input type="radio"/> Default <input type="radio"/> Top <input checked="" type="radio"/> Bottom <input type="radio"/> Starting Model

Search Settings	
Models to Consider:	<input type="radio"/> All <input checked="" type="radio"/> Loopless <input type="radio"/> Disjoint <input type="radio"/> Chain
Search Direction:	<input type="radio"/> Default <input checked="" type="radio"/> Up <input type="radio"/> Down
During Search, Sort By:	<input type="radio"/> Information <input type="radio"/> Alpha <input type="radio"/> % Correct <input type="radio"/> BIC <input checked="" type="radio"/> AIC
When Searching, Prefer:	<input checked="" type="radio"/> Larger Values <input type="radio"/> Smaller Values
Search Width: (models to keep at each level)	<input type="text" value="20"/>
Search Levels:	<input type="text" value="10"/> (leave blank to use settings from data file)

Report Settings	
In Report, Sort By:	<input type="radio"/> Information <input type="radio"/> Alpha <input type="radio"/> dDF <input type="radio"/> Level <input type="radio"/> % Correct <input type="radio"/> BIC <input checked="" type="radio"/> AIC
In Report, Sort:	<input checked="" type="radio"/> Descending <input type="radio"/> Ascending
Include in Report:	<input checked="" type="checkbox"/> H <input type="checkbox"/> dLR <input checked="" type="checkbox"/> Alpha <input type="checkbox"/> % dH(DV) <input checked="" type="checkbox"/> dAIC <input checked="" type="checkbox"/> dBIC <input type="checkbox"/> BP-based Transmission <input checked="" type="checkbox"/> Incremental Alpha <input checked="" type="checkbox"/> % Correct <input type="checkbox"/> % Coverage of Data <input type="checkbox"/> % Missing in Test

Figure 27: The “rotate” step (MBR–a2) of the bore–rotate search strategy, with search width (the number of models to be considered at each level) increased to 20.

The best AIC model was IV:ACFGJKMQRZ (level 9) with the same metrics as the slightly different level-9 model in the previous search. The wider search is preferred when deciding on feature reduction, and input variables B, D, E, H, N, P, and S were removed from the data set (by changing the mode of each of those variables to 0), and an all-model search was conducted with the remaining inputs (Figure 28).

Composition Method:	<input checked="" type="radio"/> Standard <input type="radio"/> Back Projection (Fourier)
Reference Model:	<input type="radio"/> Default <input type="radio"/> Top <input checked="" type="radio"/> Bottom <input type="radio"/> Starting Model
Search Settings	
Models to Consider:	<input checked="" type="radio"/> All <input type="radio"/> Loopless <input type="radio"/> Disjoint <input type="radio"/> Chain
Search Direction:	<input type="radio"/> Default <input checked="" type="radio"/> Up <input type="radio"/> Down
During Search, Sort By:	<input type="radio"/> Information <input type="radio"/> Alpha <input type="radio"/> % Correct <input checked="" type="radio"/> BIC <input type="radio"/> AIC
When Searching, Prefer:	<input checked="" type="radio"/> Larger Values <input type="radio"/> Smaller Values
Search Width: (models to keep at each level)	<input type="text" value="20"/>
Search Levels:	<input type="text" value="200"/> (leave blank to use settings from data file)
Report Settings	
In Report, Sort By:	<input type="radio"/> Information <input type="radio"/> Alpha <input type="radio"/> dDF <input type="radio"/> Level <input type="radio"/> % Correct <input checked="" type="radio"/> BIC <input type="radio"/> AIC
In Report, Sort:	<input type="radio"/> Descending <input checked="" type="radio"/> Ascending
Include in Report:	<input checked="" type="checkbox"/> H <input type="checkbox"/> dLR <input checked="" type="checkbox"/> Alpha <input type="checkbox"/> % dH(DV) <input checked="" type="checkbox"/> dAIC <input checked="" type="checkbox"/> dBIC
	<input type="checkbox"/> BP-based Transmission <input checked="" type="checkbox"/> Incremental Alpha
	<input checked="" type="checkbox"/> % Correct <input type="checkbox"/> % Coverage of Data <input type="checkbox"/> % Missing in Test

Figure 28: All-model BIC (not AIC) search (MBR–a3) with high width and depth for feature-reduced inputs to determine the best statistical model of the “design-all” data set.

The resulting *MBR–a model is IV:ACZ:AJKZ:CGZ:FGZ:KMZ:QZ:RZ.*

(See Section 4.2.1 for an explanation of the variable names.)

4.6.5.2 Design-All, One Output at a Time/FWD, MBR-i with Random Holdout

At the same time on a separate server, the same sequence of search types, but with Information as the criterion, were carried out for the purely information-theoretic version (MBR-i) of the search (Figure 29).

The best model by Information for this step (MBR-i1) as well as the next step (MBR-i2) was IV:ABCDGHIJKMQRZ at level 13. However, this model was rejected because it had 8191 degrees of freedom, and the model with 2047 degrees of freedom, selected also by Information (but with incremental α instead of cumulative α) was chosen. That *model is IV:ABCDGHIJKMQRZ* (level 11) with 2047 degrees of freedom.

Composition Method:	<input checked="" type="radio"/> Standard <input type="radio"/> Back Projection (Fourier)
Reference Model:	<input type="radio"/> Default <input type="radio"/> Top <input checked="" type="radio"/> Bottom <input type="radio"/> Starting Model
Search Settings	
Models to Consider:	<input type="radio"/> All <input checked="" type="radio"/> Loopless <input type="radio"/> Disjoint <input type="radio"/> Chain
Search Direction:	<input type="radio"/> Default <input checked="" type="radio"/> Up <input type="radio"/> Down
During Search, Sort By:	<input checked="" type="radio"/> Information <input type="radio"/> Alpha <input type="radio"/> % Correct <input type="radio"/> BIC <input type="radio"/> AIC
When Searching, Prefer:	<input checked="" type="radio"/> Larger Values <input type="radio"/> Smaller Values
Search Width: (models to keep at each level)	<input type="text" value="2"/>
Search Levels:	<input type="text" value="17"/> (leave blank to use settings from data file)
Report Settings	
In Report, Sort By:	<input checked="" type="radio"/> Information <input type="radio"/> Alpha <input type="radio"/> dDF <input type="radio"/> Level <input type="radio"/> % Correct <input type="radio"/> BIC <input type="radio"/> AIC
In Report, Sort:	<input type="radio"/> Descending <input checked="" type="radio"/> Ascending
Include in Report:	<input checked="" type="checkbox"/> H <input type="checkbox"/> dLR <input checked="" type="checkbox"/> Alpha <input type="checkbox"/> % dH(DV) <input checked="" type="checkbox"/> dAIC <input checked="" type="checkbox"/> dBIC
	<input type="checkbox"/> BP-based Transmission <input checked="" type="checkbox"/> Incremental Alpha
	<input checked="" type="checkbox"/> % Correct <input type="checkbox"/> % Coverage of Data <input type="checkbox"/> % Missing in Test

Figure 29: Narrow loopless-model search with Information as the criterion (MBR-i1) to determine the number of input variables (loopless levels) to be included in the subsequent “rotate” search.

This informed MBR-i3 for the removal of E, H, N, P, and S. The consequent search (Figure 30) resulted in an unexpectedly large model.

What is meant by “large” is that the model had 35 components, each consisting of three to five variables. Such a model does not appear to be parsimonious. The degrees of freedom for the model equal 65639. This indicates excessive model complexity, and means that more independent choices can be made than there are clue vectors in the entire study. However, since the model was obtained through careful and proper use of Reconstructability Analysis, it was put through the next stage of inquiry, which is the “fit” function in *Occam3*. “Fit” involves running RA in confirmatory mode (as opposed to search, which is exploratory), and is the subsequent step for deciding what model to use in prestructuring neural networks when there is any doubt about the model.

Looking through the “fit” evaluation for each component, a grouping of outcomes became apparent. There were components whose improvements by model were less than 70% (between 25 and 60 for the most part), components whose improvements by model were around the upper 90% range, and components whose improvements by model were in the negative percents.

The strategy, then, was to choose from among the components only those whose improvement values were very high. The resulting model was IV:ABCQZ: ABRZ:BFJQZ:BMQRZ:CJKZ:CKRZ:FGZ:GRZ:JKMZ:JKQZ:JKRZ: JQRZ: KMQZ:KQRZ, for which, the degrees of freedom were 65473 (down from 65639, and still appearing to be too complex for parsimonious implementation of prestructuring).

*The final MBR–i model after “fit” was determined to be ABCQZ:ABRZ:
BFJQZ:BMQRZ:CJKZ:CKRZ:FGZ:GRZ:JKMZ:JKQZ:JKRZ:JQRZ:KMQZ:
KQRZ.*

Composition Method:	<input checked="" type="radio"/> Standard <input type="radio"/> Back Projection (Fourier)
Reference Model:	<input type="radio"/> Default <input type="radio"/> Top <input checked="" type="radio"/> Bottom <input type="radio"/> Starting Model
Search Settings	
Models to Consider:	<input checked="" type="radio"/> All <input type="radio"/> Loopless <input type="radio"/> Disjoint <input type="radio"/> Chain
Search Direction:	<input type="radio"/> Default <input checked="" type="radio"/> Up <input type="radio"/> Down
During Search, Sort By:	<input checked="" type="radio"/> Information <input type="radio"/> Alpha <input type="radio"/> % Correct <input type="radio"/> BIC <input type="radio"/> AIC
When Searching, Prefer:	<input checked="" type="radio"/> Larger Values <input type="radio"/> Smaller Values
Search Width: (models to keep at each level)	<input type="text" value="20"/>
Search Levels:	<input type="text" value="200"/> (leave blank to use settings from data file)
Report Settings	
In Report, Sort By:	<input checked="" type="radio"/> Information <input type="radio"/> Alpha <input type="radio"/> dDF <input type="radio"/> Level <input type="radio"/> % Correct <input type="radio"/> BIC <input type="radio"/> AIC
In Report, Sort:	<input type="radio"/> Descending <input checked="" type="radio"/> Ascending
Include in Report:	<input checked="" type="checkbox"/> H <input type="checkbox"/> dLR <input checked="" type="checkbox"/> Alpha <input type="checkbox"/> % dH(DV) <input checked="" type="checkbox"/> dAIC <input checked="" type="checkbox"/> dBIC <input type="checkbox"/> BP-based Transmission <input checked="" type="checkbox"/> Incremental Alpha <input checked="" type="checkbox"/> % Correct <input type="checkbox"/> % Coverage of Data <input type="checkbox"/> % Missing in Test

Figure 30: Search MBR–i3 for “design-all” data that resulted in a “large” model.

4.6.5.3 Design-All, One Output at a Time/FWD, ABE–b with Random Holdout

When search ABE–b1 was carried out for the all-design data (Figure 31), the 19-component model IV:ABZ:ACZ:AKZ:ASZ:BDZ:BQZ:BRZ:CDZ:EFGZ:FJZ:HJZ:JKZ:JNZ:JQZ:JRZ:KMZ:PZ:QSZ:RSZ was found at level 37. This informed ABE–b2 to search to a depth of no more than 38, but with a width of 20. This is the “expand” portion of all-model bore–expand search.

The result of the series of searches shown in Figures 31 and 32 is the *ABE-b* model *ABZ:ACZ:AJKZ:ASZ:BDZ:EFGZ:FJZ:HJZ:JKMZ:JNZ:JRZ:PZ:QSZ:RSZ*.

Composition Method:	<input checked="" type="radio"/> Standard <input type="radio"/> Back Projection (Fourier)
Reference Model:	<input type="radio"/> Default <input type="radio"/> Top <input checked="" type="radio"/> Bottom <input type="radio"/> Starting Model
Search Settings	
Models to Consider:	<input checked="" type="radio"/> All <input type="radio"/> Loopless <input type="radio"/> Disjoint <input type="radio"/> Chain
Search Direction:	<input type="radio"/> Default <input checked="" type="radio"/> Up <input type="radio"/> Down
During Search, Sort By:	<input type="radio"/> Information <input type="radio"/> Alpha <input type="radio"/> % Correct <input checked="" type="radio"/> BIC <input type="radio"/> AIC
When Searching, Prefer:	<input checked="" type="radio"/> Larger Values <input type="radio"/> Smaller Values
Search Width: (models to keep at each level)	<input type="text" value="1"/>
Search Levels:	<input type="text" value="500"/> (leave blank to use settings from data file)
Report Settings	
In Report, Sort By:	<input type="radio"/> Information <input type="radio"/> Alpha <input type="radio"/> dDF <input type="radio"/> Level <input type="radio"/> % Correct <input checked="" type="radio"/> BIC <input type="radio"/> AIC
In Report, Sort:	<input type="radio"/> Descending <input checked="" type="radio"/> Ascending
Include in Report:	<input checked="" type="checkbox"/> H <input type="checkbox"/> dLR <input checked="" type="checkbox"/> Alpha <input type="checkbox"/> % dH(DV) <input checked="" type="checkbox"/> dAIC <input checked="" type="checkbox"/> dBIC
	<input type="checkbox"/> BP-based Transmission <input checked="" type="checkbox"/> Incremental Alpha
	<input checked="" type="checkbox"/> % Correct <input type="checkbox"/> % Coverage of Data <input type="checkbox"/> % Missing in Test

Figure 31: ABE-b1 search settings for the all-model bore-expand BIC search.

Composition Method:	<input checked="" type="radio"/> Standard	<input type="radio"/> Back Projection (Fourier)
Reference Model:	<input type="radio"/> Default	<input type="radio"/> Top <input checked="" type="radio"/> Bottom <input type="radio"/> Starting Model

Search Settings	
Models to Consider:	<input checked="" type="radio"/> All <input type="radio"/> Loopless <input type="radio"/> Disjoint <input type="radio"/> Chain
Search Direction:	<input type="radio"/> Default <input checked="" type="radio"/> Up <input type="radio"/> Down
During Search, Sort By:	<input type="radio"/> Information <input type="radio"/> Alpha <input type="radio"/> % Correct <input checked="" type="radio"/> BIC <input type="radio"/> AIC
When Searching, Prefer:	<input checked="" type="radio"/> Larger Values <input type="radio"/> Smaller Values
Search Width: (models to keep at each level)	<input type="text" value="20"/>
Search Levels:	<input type="text" value="38"/> (leave blank to use settings from data file)

Report Settings	
In Report, Sort By:	<input type="radio"/> Information <input type="radio"/> Alpha <input type="radio"/> dDF <input checked="" type="radio"/> Level <input type="radio"/> % Correct <input type="radio"/> BIC <input type="radio"/> AIC
In Report, Sort:	<input type="radio"/> Descending <input checked="" type="radio"/> Ascending
Include in Report:	<input checked="" type="checkbox"/> H <input type="checkbox"/> dLR <input checked="" type="checkbox"/> Alpha <input type="checkbox"/> % dH(DV) <input checked="" type="checkbox"/> dAIC <input checked="" type="checkbox"/> dBIC <input type="checkbox"/> BP-based Transmission <input checked="" type="checkbox"/> Incremental Alpha <input checked="" type="checkbox"/> % Correct <input type="checkbox"/> % Coverage of Data <input type="checkbox"/> % Missing in Test

Figure 32: ABE–b2, the “expand” portion of the BIC all-model bore–expand search. The result of this search is the third candidate model.

4.6.5.4 Design-All, One Output at a Time/FWD, ABE–i with Random Holdout

The initial effort for the fourth search strategy, ABE–i, crashed at the preferred depth of 500 (Figure 33), so a single-wide, 200-deep search was used instead.

Composition Method:	<input checked="" type="radio"/> Standard	<input type="radio"/> Back Projection (Fourier)
Reference Model:	<input type="radio"/> Default	<input type="radio"/> Top <input checked="" type="radio"/> Bottom <input type="radio"/> Starting Model
Search Settings		
Models to Consider:	<input checked="" type="radio"/> All	<input type="radio"/> Loopless <input type="radio"/> Disjoint <input type="radio"/> Chain
Search Direction:	<input type="radio"/> Default	<input checked="" type="radio"/> Up <input type="radio"/> Down
During Search, Sort By:	<input checked="" type="radio"/> Information	<input type="radio"/> Alpha <input type="radio"/> % Correct <input type="radio"/> BIC <input type="radio"/> AIC
When Searching, Prefer:	<input checked="" type="radio"/> Larger Values	<input type="radio"/> Smaller Values
Search Width: (models to keep at each level)	<input type="text" value="1"/>	
Search Levels:	<input type="text" value="200"/> (leave blank to use settings from data file)	
Report Settings		
In Report, Sort By:	<input checked="" type="radio"/> Information	<input type="radio"/> Alpha <input type="radio"/> dDF <input type="radio"/> Level <input type="radio"/> % Correct <input type="radio"/> BIC <input type="radio"/> AIC
In Report, Sort:	<input checked="" type="radio"/> Descending	<input type="radio"/> Ascending
Include in Report:	<input checked="" type="checkbox"/> H <input type="checkbox"/> dLR <input checked="" type="checkbox"/> Alpha <input type="checkbox"/> % dH(DV) <input checked="" type="checkbox"/> dAIC <input checked="" type="checkbox"/> dBIC <input type="checkbox"/> BP-based Transmission <input checked="" type="checkbox"/> Incremental Alpha <input checked="" type="checkbox"/> % Correct <input type="checkbox"/> % Coverage of Data <input type="checkbox"/> % Missing in Test	

Figure 33: ABE-i1 set at 200-deep after the 500-deep search crashed with no results after several days.

The candidate model using overall α was excessively large, with 76 three- or four-input terms and 200 degrees of freedom. In comparison, the incremental- α model had 32 terms and 57 degrees of freedom. This was selected as the *ABE-i candidate*:
ABZ: ACZ:AGZ:AKZ:AMZ:ARZ:ASZ:BDZ:BNZ:BQZ:BRZ:CDZ:EFGZ:FJZ: FQZ:FRZ:GHZ:GPZ:GQZ:GSZ:HJZ:HPZ:HQZ:HSZ:JKRZ:JNZ:JQZ:KMZ: MQZ: NQZ:PQSZ:RSZ.

4.6.5.5 Summary of the Four Searches

The purpose of these four searches was to identify four candidate models with which to prestructure neural nets, whose performance metrics (based on a required number of repetitions) are then compared with those of the fully connected networks,

RA performance, the human benchmark, and the “lousy” control network. Out of the four candidates, two finalists were chosen (based on generalization-performance improvement with the initial neural-net runs evaluated with the holdout set. These finalist networks are based on the two RA models ***ABZ:ACZ:AJKZ:ASZ:BDZ:EFGZ:FJZ:HJZ:JKMZ:JNZ:JRZ:PZ:QSZ:RSZ*** (ABE–b) and ***ABZ:ACZ:AGZ:AKZ:AMZ:ARZ:ASZ:BDZ:BNZ:BQZ:BRZ:CDZ:EFGZ:FJZ:FQZ:FRZ:GHZ:GPZ:GQZ:GSZ:HJZ:HPZ:HQZ:HSZ:JKRZ:JNZ:JQZ:KMZ:MQZ:NQZ:PQSZ:RSZ*** (ABE–i).

CHAPTER V. APPLICATION AREAS

5.1 Neural-Network Prestructuring

As a technology that has been successful in unexpected areas and challenging applications, such as machine pronunciation of English [14, pp. 1–3 and 182–185; 157, p. 10], the field of Neural Networks is one of the key players behind the scenes in much of today’s advanced technology in the adaptive implementation of control, filtering, decision making, and pattern recognition. Neural nets are used where decision-making capabilities akin to human intuition are needed, and the data are too numerous or multi-dimensional for humans to handle. Although crowd-sourcing is increasingly used in the physical sciences (astronomy and environmental science, for example), its performance is unpredictable, inconsistent, and even less thoroughly studied than those of Computational Intelligence methods. Neural networks are among the few technological solutions that offer a reliable way to make intelligent inference using today’s overly abundant data.

At the same time, the difficulty of interpreting *why* a neural network performs a certain classification (or makes a control decision, etc.), along with the highly heuristic approach to their design detracts from the potential for widespread use. This dissertation is one of many efforts to systematize some aspect of neural-network design, in this case focusing on the interconnection structure. The interconnection structure of a network is a tangible, visibly representable aspect of neural-net design that (as opposed to learning-rate schedules, activation functions, or entropy values [85]) designers, users, or project managers who do not have a thorough understanding of the

field can still understand and work with. Hence, the design methodology of discovering problem-domain structure through Reconstructability Analysis and translating that structure directly (visually) to a neural-net connection scheme has the advantage that even non-specialists working with neural networks can understand and implement it.

In this dissertation, Reconstructability Analysis (RA) models are directly translated into neural-network interconnections. As an example, let's take the RA model from the ABE–b search: ABZ:ACZ:AJKZ:ASZ:BDZ:EFGZ:FJZ:HJZ:JKMZ:JNZ:JRZ:PZ:QSZ:RSZ. The network structure shown in Figure 34 is a direct mapping from the inputs A –S to the 14 hidden-layer elements. The bias and its connections have been removed to make the model-based connections easier to discern. Each hidden-layer element's inputs correspond to one of the terms in the RA model.

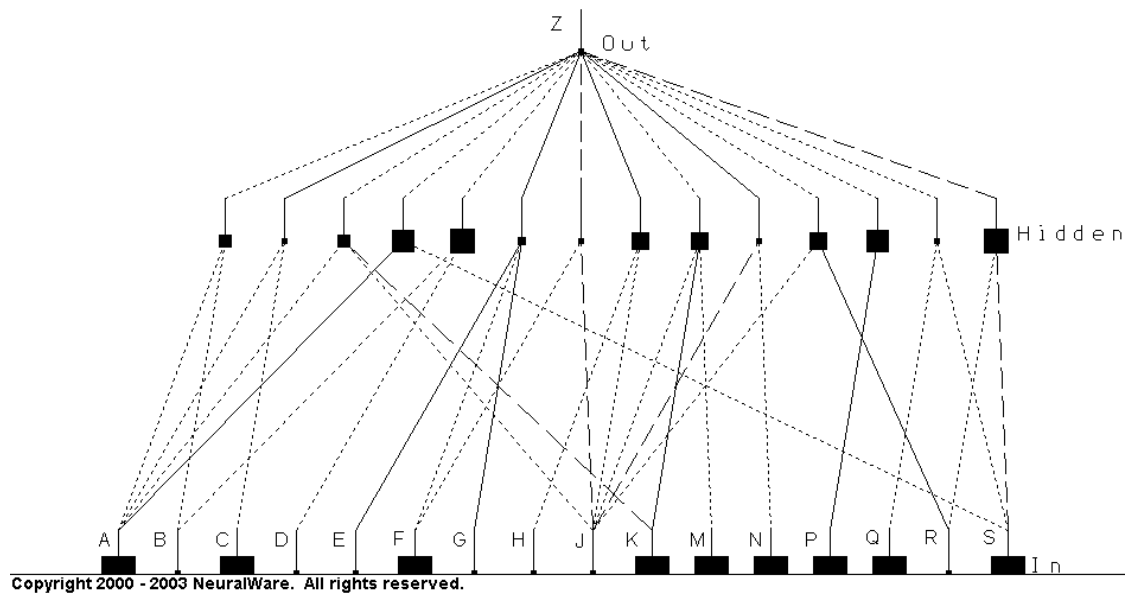


Figure 34: Prestructured neural network (multilayer perceptron) based on the BIC-based model resulting from the all-model borexpend (ABE) search.

5.2 Music Training

What follows is an almost identical version of the paper that has been published in the Journal of Music, Technology and Education, Vol. 4, No. 1, included for the purpose of highlighting the potential application of the clave-direction aspect of the present research in music technology. The purpose of this article is to introduce clave-direction recognition⁸⁵ (as opposed to detection of the clave rhythm or the clave instrument) for generalized rhythm patterns, explain its significance and potential role in music education and music-making, propose a system-level approach for its implementation and to suggest application areas with the potential to benefit performers, educators, students and the music industry.

5.2.1 The Growing Relevance of Clave

In the current age of globalization and the Internet, with its consequent worldwide dissemination of virtually all forms of traditional and modern music,

⁸⁵ The word ‘clave’ comes from the Spanish language. It means ‘key’ or ‘legend’ as in the legend for a map or the key to a code. In rhythmic musical usage, it can take on multiple meanings. One is clave the instrument (las claves). Another is clave in the harmonic sense (meaning ‘musical key’ in Spanish as in ‘the key of C#’). A third meaning is the class of standard accent patterns such as son clave and rumba clave, referred to here as narrow-sense clave. The fourth and final meaning is clave as an organizational principle: wide-sense clave. In the wide sense, clave recognition is not simply noting the existence of particular patterns (although that is a non-trivial task), but of the relationships that any arbitrary pattern can have with a family of related traditional patterns. This is because clave is not just a pattern but a temporal guide: “The clave is not a beat, as we understand beats in North American music. [...] The clave is a key: a way of coordinating independent parts of a polyrhythmic texture. [...] There are any number of rhythmic formulas played by various instruments that are images of the clave, any of which is sufficient to tell the other musicians where the rhythmic key is.” [9, pp. 170–71]

In the same vein, Spiro states that it is “the ‘key’ that determines how the complex rhythms and syncopations of African-based music are to be assembled, arranged, performed and even improvised” [158, p. 12] and Rodriguez maintains that “Clave is inherent in *all* [original emphasis] aspects of Afro-Cuban music [...] from rhythm to melody” [8, p. 41].

geographic isolation is no longer an obstacle to social contact, cultural exposure and musical collaboration. As a result, musicians have been paying greater attention to the cultural norms and musical contexts of various traditions from which they borrow artistic elements. Hence, acquiring and exercising traditional knowledge by individuals who are not members of a host ethnic group is becoming a common prerequisite to creating culturally sanctioned musical fusions and authentic renditions of folkloric music. One of the ways in which technology has the potential to come to the aid of music education is the rapidly growing field of Music Information Research (or Music Information Retrieval (MIR)). The literature in this field abounds with work concerning the recognition of style, genre or singer, the detection of beats and onsets, the induction of tempo, meter and key and other aspects like mood and structure. (References to some of the peer-reviewed articles on a subset of these topics relevant to clave detection were listed in the appendix at the end of the journal article, but have been removed from the dissertation copy. The interested reader can find the citation for the full article at the end of this chapter.)

However, even with all these developments in MIR technology, some important musical principles have barely been explored. One of these is the temporal organizational principle of West African-based Latin American music (henceforth referred to as ‘Afro-Latin music’). One commonly recognized name for this principle is *clave* [1, p. 324; 2, p. 73; 3, p. 5; 5, p. 10; 7, p. 1; 8, p. 41; 11, p. 3; 158, p. 12; 159, p. 647; 160, p. 60]. Much of the music of Latin America revolves around this subtle rhythmic structure that governs the ebb and flow of rhythmic tension in the form of relative

offbeatness [54, pp. 18–27] as a function of time. This structure is not solely for percussion instruments but manifests itself in salsa piano, reggae vocals, bossa nova guitar, *axé* horn arrangements, and in countless other ways. Clave represents a sense of *dynamic movement and concord* based on relative *offbeatness*. (This particular use of the term ‘clave’ is not without controversy, but is still widely accepted.)

A working knowledge of clave – clave consciousness – is difficult to acquire for those who have not been brought up in a clave-based cultural environment. Technological aids that guide composing, arranging, music production and training are therefore needed for a growing population of non-native musicians who wish to learn to function within the idiom.

5.2.2 A Note on Terminology

There is much controversy over the term clave when applied to music that is not from Cuba. Specifically, in the case of Brazilian traditional music, there is a popular claim that clave does not apply [161, p. 88; 162]. On the contrary, such folkloric Brazilian styles as samba and *maracatú*, and modern popular music like *axé*, operate firmly within the *Afro-Latin concept* of clave. Claims to the contrary seem to have more to do with concerns about cultural misunderstandings than any conceptual difference regarding the common West African roots of the music⁸⁶. Although the term appears to have originated in Cuba [9, pp. 93–96], it has come to be widely used in the United

⁸⁶ The primary concern for Brazilian musicians seems to be the general lack of recognition that the language of Brazil is Portuguese, not Spanish. For this reason, the use of the Spanish word clave for Brazilian music can be off-putting. In addition, because of the subtle differences in how the clave concept works in Cuba and Brazil, it is simpler sometimes to claim clave does not apply—because one typically associates clave with Cuban clave.

States. However, in many clave cultures themselves there is no commonly accepted name and little or no explicit analysis of this construct. Nevertheless, the concept – with or without a particular label – plays a central role in the temporal organization of Brazilian, Haitian, Trinidadian, Belizean, Ghanaian, Nigerian and other African-Diasporan musics just as it does in Cuban. Its central role in Brazilian music is evidenced in the stern disapproval by Brazilian masters of those who unknowingly break the rules of clave (cross clave).

The principle underlying the clave-direction concept is evident in a variety of rhythmic patterns beyond the basic five-note sequences (which may be found in [2, p. 75; 8, p. 41–45; 11, p. 157–68; and 158, p. 13] commonly associated with the term *clave*). Recently, a small yet excellent body of MIR research has developed that addresses computational analyses of clave patterns such as Toussaint [11, 54]; Wright et al. [159] and Flanagan [163]. These and similar studies are focused either on developing metrics of complexity (such as mathematical and musical measures of difficulty and ambiguity) or on the shift-invariant differentiation of the various related sequences from one another. In contrast, here the common underlying principle of temporal harmony (‘harmoniousness’ that is not specific to pitch) that concerns all possible accent patterns, whether they consist of five onsets or not, is discussed⁸⁷. A common example of such a pattern is given in Figure 35.

⁸⁷ While Toussaint and colleagues are interested in differentiating 3-2 son clave from 3-2 bossa clave in terms of mathematical and psychological complexity, and Tzanetakis and colleagues are interested in the shift-invariant recognition of any 2-3 or 3-2 (or otherwise shifted) son clave, the present article’s point of view is the need to recognize what 2–3 bossa, 2–3 son, 2–3 partido alto, 2-3 cáscara and a potentially infinite number of other 2-3 rhythms have in common that can be extracted,

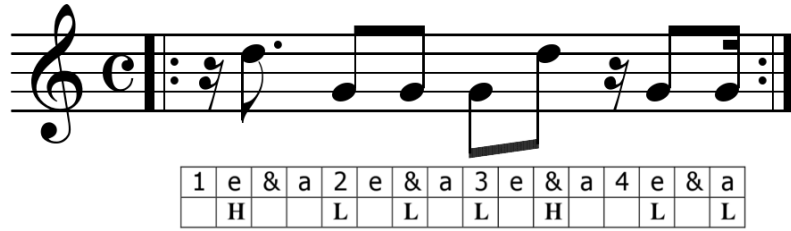


Figure 35: *Partido-alto*, the rhythmic basis of samba, in 3-2 clave direction.

In many other countries, and among scholars worldwide, various other names and descriptions are in use to refer to the same fundamental principle. These include *clips* [164, p. 31] and *compás* [165], and in Uruguayan *Candombe*, both clave and *madera* [5, pp. 10, 13, 22]. Further descriptive phrases favored by scholars to describe the duple- and triple-time clave-type patterns include “the principle of mobility and finality” [103, p. 173], “generative concept” [158, p. 7], “rhythmic key” [9, p. 168], “musical delimiter” and “timeline” [116, p. 61], “standard pattern” [12, pp. 116–121; 54, p. 22], “structural device” [7, p. 5], and “the key pattern” [12, p. 160–165]. Novotney also lists eighteen other names and descriptions for compound-time clave-type patterns [12, p. 168].

It is due to this proliferation of labels and descriptive phrases, as well as the familiarity of the word clave due to its use in popular Cuban music, that the term is preferred here despite the associated controversy.

understood and utilized, and how this differs from all 3-2 patterns or from a complete lack of clave direction. These two directions of research are orthogonal. Rather than the shift-invariant recognition of a particular clave example, the approach promoted here concerns phase-dependent relationships.

5.2.3 Introduction to Clave

While commonly described as one of several five-note percussion patterns, clave (Figure 36) actually represents more than a rhythmic phrase or even a set of related rhythmic phrases. It goes beyond percussion, establishing rhythmic ground rules for all voices and instruments in the Afro-Latin idiom [2, pp. 74, 75; 3, pp. 5, 7, 10; 8, p. 41; 158, p. 12]. This is because the well-known rhythmic phrases (exemplified in Figure 36, and in [2, pp. 26, 73, 75; 5, pp. 10, 13; 7, pp. 3, 6–8; 8, p. 43; 158, pp. 12, 13]) are indicative of an underlying system of rhythmic harmony, similar to the familiar systems of tonal harmony that govern the simultaneous (vertical) and progressive (horizontal) aspects of pitch relationships found in musical traditions throughout the world.

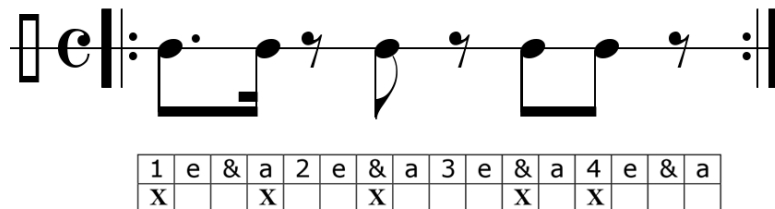


Figure 36: The son clave: the Cuban name for a common pattern in Brazilian, Caribbean and West African music, as well as early rock'n'roll and jazz. It is the best-known of all clave patterns.

5.2.4 Systems of Musical Harmony

Although theories and traditions of music differ from one culture to another, by and large, it is expected that some convention be present to govern the progression of the pitch aspect of music. Examples of such systems include the Indian *raag*, the Turkish *makam*, the Baroque figured bass and various standard chord progressions like

the twelve-bar Blues form, ‘I Got Rhythm’ changes and the distinct harmonic sequences of the *samba de enredo*. It is much less common, however, to find an established body of rules that governs the simultaneous and progressive aspects of rhythm (attacks, durations and accents) outside the African Diaspora and the Indian Subcontinent. Clave, as it has been interpreted in the New World over the course of the past six centuries, is such a system of temporal non-pitch harmony, governing both the simultaneous execution of rhythmic attacks and releases (not just percussive parts) and the ebb and flow of rhythmic tension within each phrase.

5.2.5 Clave as a Principle of *Rhythmic* Harmony

How does this system work? What is meant by clave consciousness? This is not merely the awareness that a clave pattern is present – although that in itself is a challenge to many listeners. At times, clave-based music may exhibit no explicit clave pattern at all. The presence of the sense of clave in this type of music is sometimes called ‘implied clave’. Musicians who are clave-conscious do not necessarily play the clave pattern. In fact, they do so quite rarely, unless they happen to be playing specific instruments in specific settings⁸⁸. Instead, musicians are informed by the clave, which may just be a mental pattern. They hint at it, draw from it for improvisation, depart from it for tension and return to it for resolution⁸⁹. In his piano book, Herder calls clave a ‘demanding and inflexible rule’ and refers to it as ‘the law of clave’ [3, p. 5]. He also

⁸⁸ Examples include the *caixa* in *samba batucada*, the *repique* in *samba-reggae* and the instrument called *claves* in *rumba guaguancó*.

⁸⁹ “Improvisation depends [...] on thinkers having absorbed a broad base of musical knowledge, including myriad conventions” [166]. Clave is one such convention.

distinguishes the clave *pattern* itself from this ‘law’, a distinction presented here as narrow-sense vs. wide-sense clave. It is thus understood that there are ‘correct’ and ‘incorrect’ ways of playing within the clave idiom. Clearly, there are no moral or legal ramifications for playing incorrectly, but the musician who does so will generally find they have interfered with the flow of the music and that the audience has difficulty feeling the groove. Mauleón-Santana (author of one of the clearest explanations of clave and its function) concurs in her piano book: ‘all of the rhythmic, melodic and even harmonic elements within the structure of an arrangement must be correctly aligned with the clave [...] as [must] the possibilities for improvisation’ [7, p. 8].

Playing the ‘incorrect’ way is called ‘crossing clave’. One can, of course, *choose* to break with convention – especially if one knows the rules thoroughly or if one is functioning clearly outside the idiom. When executed and resolved skillfully, crossing clave is a common and acceptable way to build tension and ‘add spice’ to the music. Outside of such conscious artistic prerogative, however, crossing clave is an Afro-Latin music taboo. The term ‘taboo’ is not meant in an absolute sense of wrong; it is merely intended as an indication of the potency of convention within the idiom. It also has no bearing on rock or jazz performance (as long as a strong connection to traditional Afro-Latin music is not claimed). Like Herder and the present author, Mauleón-Santana differentiates the instrument, the pattern and the concept of clave [7, p. 1], arguing in no uncertain terms that clave is a rhythmic, melodic and harmonic rule [7, pp. 7–9], ‘albeit a very vague one’ [7, p. 9], and adding that it is subject to artistically justified violation [7, pp. 8–9].

5.2.6 A Closer Look at Some Intricacies of Clave

As stated above, before focused enquiry, clave appears to be no more than predefined patterns of rhythmic accents with two distinct and discrete halves. This, however, is an oversimplification of the cognitive and organizational role that clave plays [2, p. 74; 7, p. 8] in Afro-Latin music. As a result of this oversimplification, students of clave face a choice between blindly accepting aspects of clave relationships that do not follow from this simple explanation and trying to make sense of clave by coming up with special cases or exceptions to explain the perceived inconsistencies⁹⁰. For instance, if what makes 3-2 clave what it is are the three note onsets in its first half and the two note onsets in its second half, how can the *cáscara*, the *partido-alto* or the *son montuno* (each with seven or more onsets per clave phrase) be said to ‘be in clave’? Similar problems arise in the following onset patterns (Figures 37–46), presented in both standard European notation and a simple variant of Koetting and Harland’s time unit box system (TUBS) notation [158, pp. 115–46]:

1	e	&	a	2	e	&	a	3	e	&	a	4	e	&	a
				X	X		X					X	X	X	

Figure 37: A possible samba *surdo* pattern in 2-3 clave direction: Comparing numbers of onsets in each half is evidently not helpful.

⁹⁰ The ‘third surdo’ in *samba carioca* is a typical example of a confusing, seemingly clave-contradicting, pattern. The author has known performers and teachers of Afro-Brazilian music refer to it as an exception, whereas the author shows in an upcoming music-theoretic article (submitted to *Current Musicology*) that this is not at all an exception, but solidly founded in the Brazilian interpretation of clave.

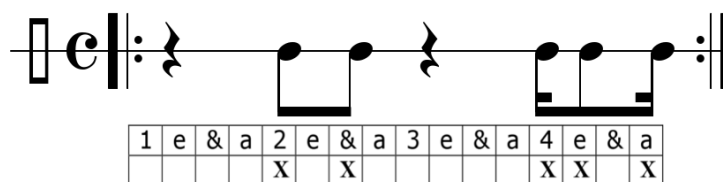


Figure 38: One of the standard *surdo* patterns in Rio-style samba employed by G.R.E.S. Mocidade Independente de Padre Miguel and many other samba schools. (The notation disregards samba swing.) This pattern matches the 3-2 *bossa* clave or is “in 3-2.” Note that there are two onsets in the first half and three in the second. This contradicts the simple-minded approach of counting onsets in clave halves, a practice that possibly owes its popularity to the ubiquitousness of the *son* clave. On the other hand, reconciling this with the idea that this pattern is in 3-2 clave orientation is straightforward and free from contradiction once one becomes familiar with the relative nature of clave offbeatness (Appendix A, submitted).

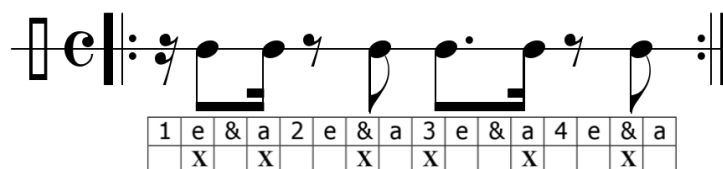


Figure 39: A common accent pattern that can be heard in both traditional and modern Afro-Brazilian music, this pattern is (perhaps surprisingly) typically associated with the 3-2 clave direction even though the second half of the pattern is identical to the 3-side of both *son* and *bossa* clave.

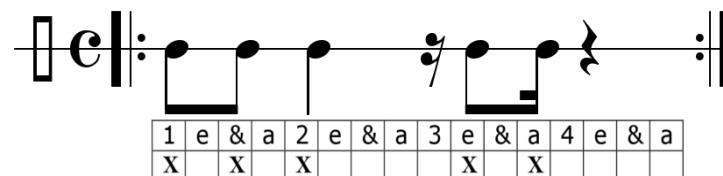


Figure 40: A contrived (not traditional, but still likely) pattern of accents in 2-3 clave orientation: Once again, note that the first half has three onsets while the second half has two. Yet, the overall pattern is in 2-3, not 3-2. This is similar to the point made in Figure 9 above and Figure Twelve in Appendix A.

1	e	&	a	2	e	&	a	3	e	&	a	4	e	&	a
	H		L		L	L		H	H		L		H	H	

Figure 41: A typical 3-2 agogô (bell) pattern in Rio-style samba: Yet again, counting the onsets (even if one uses the foreshadowing concept) is of no help in determining clave orientation.

1	e	&	a	2	e	&	a	3	e	&	a	4	e	&	a
H		L		L	L			H	H		L		H	H	

Figure 42: The 3-2 agogô pattern of Figure 41 becomes 2-3 when phase-shifted by one onset.

1	e	&	a	2	e	&	a	3	e	&	a	4	e	&	a
X		X		X		X		X		X		X		X	

Figure 43: The accent pattern (ignoring grace notes) for the caixas (snare drums) in the Rio samba school G.R.E.S. Estácio De Sa [167, p. 51]. Since the only note breaking the steady pattern of eighth notes (onbeats) is the last one, that is the only clave-indicating onset, and the pattern is in keeping with 3-2 samba direction.

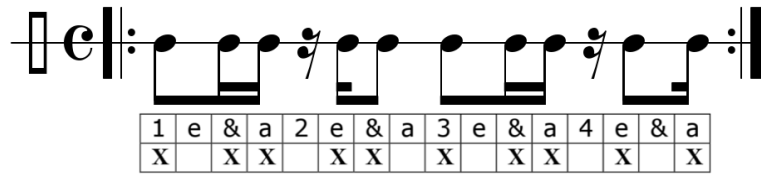


Figure 44: One possible notational abstraction of the standard caixa pattern in the Mangueira samba school: This pattern matches the 3-2 patterns in the rest of the ensemble. The only clue as to the sense of clave direction is in the difference between the offbeat sections following the second and fourth downbeats ([2]-E-AND vs. [4]-E-A). This is how clave is subtle: A single 16th-note shift, i.e., the placement of one onset (out of ten), is enough to set the clave direction.

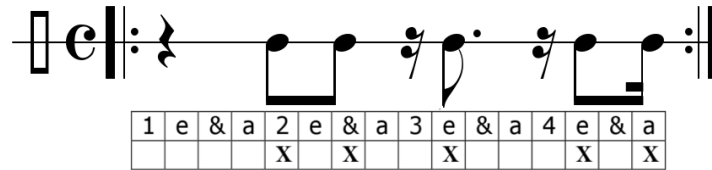


Figure 45: The second bar of a two-bar phrase used by the third *surdo* in the Rio samba school Império Serrano [167, p. 53], this is a good indication of the subtlety and relativity of samba offbeatness. A method of cancellation of like onsets in corresponding portions (similar to cancelling off terms in algebra) yields the pattern of Figure 46, which is clearly in 3-2, while the original pattern shown here initially appears to establish no clave direction.

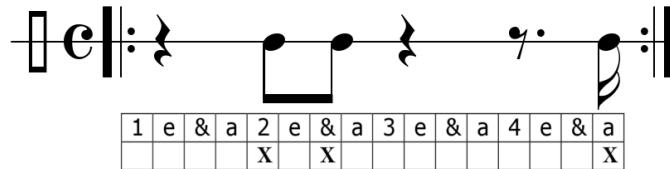


Figure 46: The pattern of Figure 45 after cancellation, which results in a pattern closely related to the standard 3-2 *surdo* pattern in Figure 38.

These examples reveal two aspects of clave-direction discernment that support the case for ‘intelligent’ technological tools: (1) Even with notation and not having to be executed, heard and discerned in real time, clave-direction recognition is a challenging

task. (2) There is a structure and consistency that can be exploited to make automating this task achievable.

5.2.7 Why is the Study of Clave Important?

Keepers of the tradition consider it important for everyone who is making or teaching Afro-Latin music to be able to function within the (orally kept) rules of the music at a level of proficiency close to that of practitioners who have been brought up immersed in these cultural subtleties. This calls for a thorough understanding of the rules.

Henderson argued that ‘accurate and complete criticism of a fugue is only possible to one who is fully acquainted with the laws of fugue’ [168, pp. 33–34]. Likewise, thorough understanding of the Cuban *guaguancó*, for example, is not possible without being fully acquainted with the laws of the *rumba*, which include the Cuban concept of clave. However, neither counting the numbers of onsets nor considering the first and second halves of the phrase is necessarily helpful in determining whether a pattern follows a particular clave direction (Figures 37–46). The patterns of Figures 37–46 hardly even resemble any of the standard clave forms, but they each possess an implied clave direction nonetheless: they are neither clave-neutral nor completely outside the concept of clave. It is imperative, therefore, to attain clave consciousness (if playing clave-based music) either through tradition and upbringing or through conscious effort.

5.3. Clave and Music Technology

5.3.1 Why a High-Tech Solution to Clave Training?

How, then, does one know the right way to play? This is where technology can be of assistance. Advanced signal analysis and the modeling of complex human cognition via Computational Intelligence (CI) can enable the development of clave-recognition, clave-analysis and clave-teaching tools for musicians, listeners, researchers, music-collection managers, students and teachers. The typical *traditional* mode of learning involves years of exposure beginning very early in life. Some argue that this training begins in the womb, as pregnant mothers do the samba, *yambú*, *comparsa* or other locally ubiquitous dance. However, at least one psychological study on rhythm suggests that rhythmic awareness arises *after* the first year of life [169, pp. 199–200]. Either way, early exposure appears to be crucial in clave enculturation. This type of exposure can hardly be duplicated by people from outside the culture – yet outsiders wish to play this music, and to do so while respecting the traditions. Musicians who have clave-based music as part of their culture often find it difficult to explain why clave works the way it does. It is not uncommon to hear that ‘it is in the blood’. However, it would be scientifically naïve (not to mention, offensive) to say that blood (or rather, genes) alone can carry some innate analytical ability for understanding and following specific musical organizational schemes. A lifetime of shared culture is a more likely explanation for clave consciousness. In contrast, when a comparable period of time is not available for learning, electronic devices that deliver distilled expert knowledge to end-users through state-of-the-art digital signal processing (DSP) and CI techniques

would make excellent companions to accelerated but intensive study. As an analogy, consider that the compound time signatures of Balkan folk music come naturally to non-musical natives, but baffle even experienced drummers from other backgrounds. Similarly, the swing feel of a be-bop drummer is cultivated over a lifetime of listening and performing, just as interaction between the lead drummer and the dancer in a Wolof, Yorùbá or Asante ceremony requires the command of a sophisticated artistic vocabulary. In the same way, a working knowledge of clave (a feel for clave) takes a comparable period of study and exposure to cultivate. This is true even for professional musicians. It takes dedication, immersion and even some familiarity with religious, linguistic and historical elements to develop a feel for clave in any one of its geo-cultural contexts (such as in Cuban, Ghanaian or Brazilian music). Closing the gap can be made easier with technological aids that guide composing, arranging, music production and training for a growing population of nonnative musicians learning and performing Afro-Latin music. A high-tech solution is proposed here because the simplicity of clave is deceptive. A comparison of Figure 47 (waveform for 3-2 son clave) with any readily available audio waveform or sheet music will attest to this apparent simplicity. In its basic form, virtually anyone can recognize it, and most musicians can play it. However, recognizing how and why the notion of clave influences virtually every temporal aspect of a composition or an improvisation, and to be accurately guided by it when creating music, are far from simple. They constitute a multifaceted, intricate cognitive task that trained individuals can perform with ease, in much the same way a baseball player can strike a ball with great accuracy – something that requires a set of physics calculations

no computer can yet execute in real time. Similarly, the recognition and idiomatically accurate use of clave is a complex mathematical and psychoacoustical task that one can learn only after extensive exposure and practice – a process that can be accelerated and assisted by technological clave-training devices.

Once developed, clave-aware training and recording technology could be put to use in audio recording, notation and sequencing, auto-accompaniment keyboards, music query systems and various types of electronic music coaches, existing or to be developed.

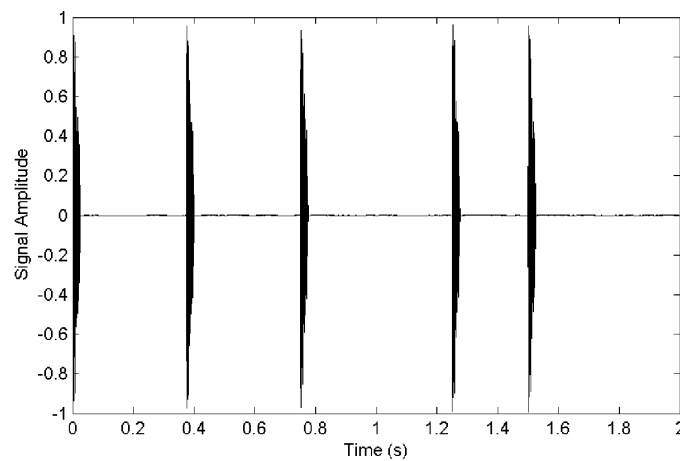


Figure 47: The audio waveform for one bar of 3-2 son clave, as captured in MATLAB.

5.3.2 Current Research and Appropriate Technologies for Machine Recognition of Clave Direction

The implementation of clave-aware systems requires the integration of many aspects of music technology into a *machine listening* system. These aspects include source separation, beat detection, meter identification, tempo extraction, quantization and

various musically informed decision processes regarding timing fluctuations and dynamics. These are needed to obtain attack-point rhythm (or some other appropriate abstraction) from audio sources. Once a sufficiently simple representation is obtained for each audio stream, the author’s present research in neural networks suggests that machine recognition of clave direction is attainable with great accuracy for a given cultural context. The proposed representation for basic clave-direction recognition is attack-point rhythm, i.e. time-quantized and dynamics-quantized rhythm with durations and pitch information removed. This abstraction is necessary for the initial development of mimicking clave consciousness in machines; subtler musical aspects (duration, dynamics and pitch) that are relevant to clave direction can be included in later designs. (Hence, both theoretical work and commercial solutions will eventually need to address these factors.)

As an example of the highly complex psychoacoustical processes humans engage in for determining clave direction, consider the *a cappella* introduction to the Daniela Mercury song *Nobre Vagabundo* from the album *Feijão Com Arroz* [170]. This vocal segment can be reduced (with some difficulty!) to the following attack-point representation: 1001001000000101 | 1001010101000000 | 1011011000010001 | 1101101000100000 | 1101011000100101 | 1001011101000000 | 1011011000000101 | 1011011000100000⁹¹. The overall clave direction is clearly 3-2. Such a conversion is currently not within technological capability in the complete absence of contextual information or human-operator guidance and even requires years of training for humans

⁹¹ These are the onsets for “Quanto tempo tenho, prá matar essa saudade; Meu bem o ciúme, é pura vaidade; Se tu foges o tempo, logo traz ansiedade; Respirar o amor, aspirando liberdade” [170].

to carry out. However, as musical signal-analysis techniques mature and generalize beyond the narrow constraints of most current demonstrations (solo piano only, percussion only, European folk songs only, electronic dance music only, score-guided, etc.), the high-level framework described in the next section will become feasible.

5.3.3 A high-level framework for realizing automated recognition of clave direction

The complete set of technologies to implement ‘blind’ clave-direction recognition would involve the integration of the following stages (some in parallel): decompression (if necessary), genre detection, source separation, beat detection (because clave direction is primarily concerned with note attacks), meter induction, contextual decision making⁹², quantization, conversion to vector form for Machine Learning and the final stage of inductive and intelligent discernment of clave direction.

Different data types and sources require different combinations of these actions. From the simplest to the most challenging, four potential types of data that may be encountered are (1) annotated archive data (MIDI, MPEG-7, XML, etc.), (2) real-time annotated data (MIDI stream), (3) real-time audio and (4) compressed audio (such as MP3). For the annotated cases, a relatively simple programming stage can precede intelligent clave-direction recognition. For the audio cases, a set of music-specific DSP

⁹² Artists like Chico Science, Lenine and Airto have created music-in-clave in 7/8. Nonetheless, odd or compound meters constitute a miniscule portion of the repertory. On the other hand, much Latin American and West African sacred music is in compound (triple) time. The literature on the connections between duple- and triple-time clave-based musics is in its infancy, with only a handful of sources on the history (not so much the musicology) of the transition from the West African triple-time roots to the Latin American duple-time synthesis of today. At this early stage of development, applying automated recognition of clave direction only to music in meters convertible to common or cut time is sufficiently ambitious.

algorithms (see the DSP and MIR technologies in the appendix) precede the intelligent recognition phase (and are sometimes called ‘machine listening’).

With annotated data, information about the musical context, like tempo, meter and quantization, can be directly accessed from the file header. With a known meter and tempo map, conversion from a MIDI event list to an attack-point vector is a trivial programming task⁹³. When the annotated data must be processed in real time, a separate (parallel) clave-recognition chain will be needed per channel. This is because clave-based music typically includes one or more layers of clave-neutral ‘grounding’ rhythm parts (such as the hand claps in *samba de roda*, the *okónkolo* in *batá* music, *surdo de primeira* in *samba batucada* and the *martillo*), and clave-based or clave-determining parts (such as the *gã* bell in *samba de roda*, the *iyá* in *batá*, *surdo de terceira* in *samba batucada* and the *cáscara*).

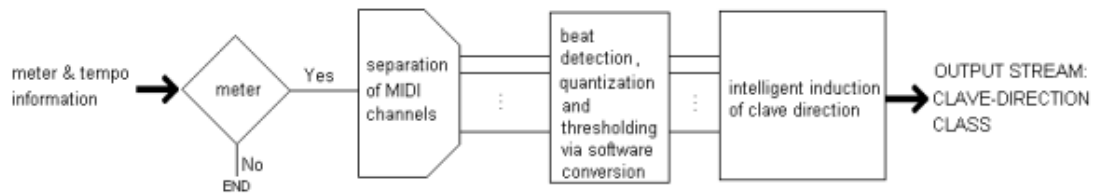


Figure 48: A system-level data-processing stream (with human interaction) for identifying clave direction in MIDI data received in real time.

⁹³ The musical-contextual challenges are in the choice of an amplitude threshold for differentiating between clave-relevant attacks and ghost notes or embellishments and in the choice of a time window for interpreting expressive timing. Additional research is required into the effect and role of expressive timing within the clave concept—the present research disregards tuplets, flams and swing, but microtiming studies do exist.

Each channel's header is read to accept or reject the musical selection based on its attributes. If accepted, each stream must be quantized, converting its events to a series (vector) of attack points. These vectors, in turn, are classified into their clave directions.

With actual audio, the challenge shifts from the music-theoretic to the technological. General-purpose recognition of many aspects of music, which is difficult enough for humans, will need to be executed, possibly in the absence of contextual information from a human attendant. To date, successful demonstrations of MIR techniques for beat detection, meter induction and the like have required strong constraints on the input stream. To expect general-purpose machine listening to work in real time is reasonable. To expect it to work on the first pass is also fair. To expect it to work without human input of contextual/categorical information may also become feasible in the near future. (However, to expect it to work in real time on the first pass without human contextual input is probably not.) As for the degree of simplicity or ambitiousness of the clave interpretation, all of the above can be implemented in a simple fixed manner or to an arbitrary degree of adaptive sophistication. For instance, thresholding could be realized with a single fixed ratio below which note events are seen as clave-irrelevant or it could be designed with intelligently adaptive algorithms that take cues from the dynamic variations observed in the audio stream for more human-like discernment of relevant note attacks.

5.4. Clave Analysis as a Music-Production Feature

5.4.1 Music-Making

In a typical, but simple music-making scenario, one would specify the overall (or initial) clave direction for a piece being composed, recorded or sequenced just as one specifies a tempo, key and meter. A clave-analysis tool could then flag occurrences in the recorded, notated or sequenced music that go counter to the chosen clave direction (cross clave). It may even be possible (in advanced products) to set a ‘vigilance’ or ‘strictness’ parameter or a cultural-context variable. It is then up to the musician to take into account or to disregard this information.

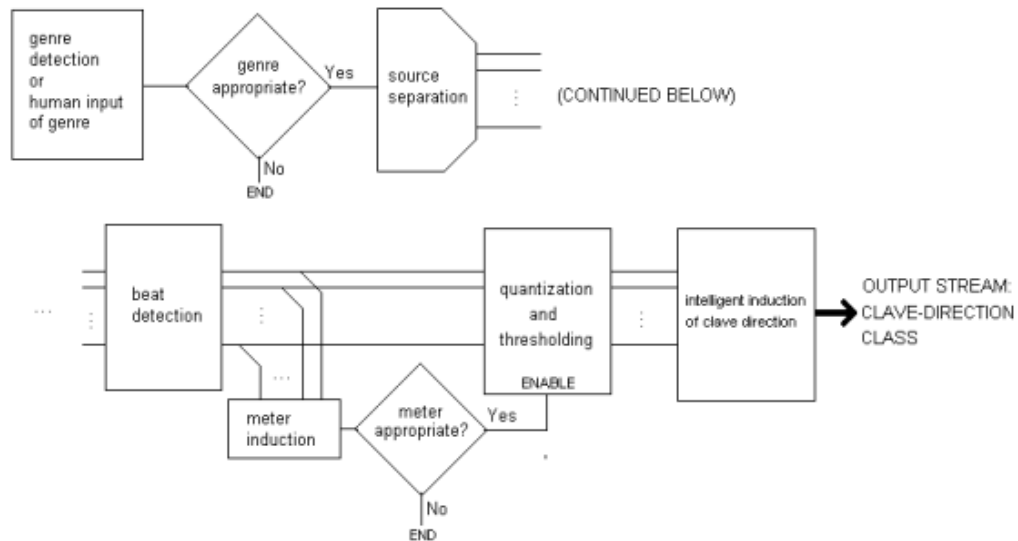


Figure 49: A system-level data-processing stream (with or without human interaction) for identifying clave direction in multi-channel recorded or multi-instrument live audio.

This issue of clave-aware editing in computer-aided music production remains unaddressed even as the music industry develops ever more Latin-music-oriented

products. A number of recently released metronomes scratch the surface of the need for clave-training tools for musicians by incorporating standard five-note clave patterns among their click options, but electronic drum trainers and most other intelligent or interactive music products have yet to offer *clave-concept* features. Looking at clave from the point of view of notation, MIDI or digital-recording applications, it is natural to expect advanced music equipment to rule-check MIDI or audio files for a selected clave direction. Setting one or more tempi, time signatures and keys is part of the nature of making music. Within the Latin American idiom, it is equally natural to set a clave direction (either universal to the piece or subject to change), and to expect advanced music software to be able to flag deviations from it. It is up to the artist to make the choice of accepting or rejecting the suggested corrections, as not all musicians will aim to function squarely within the confines of strict clave convention. Those who already know how to function within the tradition may consciously choose to ‘cross clave’ to create contrast and tension in composition or improvisation. Those who are not sufficiently familiar with the workings of clave can use this type of feedback to improve their art. In either case, having the option to receive clave-based feedback is a benefit with no downside.

5.4.2 Music Training

In the last decade, the music industry has developed and marketed a small number of ingenious products for the percussion-training market, such as the Roland RMP series of drum trainers, which give the user feedback on stroke evenness and timing fluctuations. Products like these are ripe for the inclusion of ‘clave

consciousness' (in the form of software modules) because they already implement prerequisite features like beat detection and built-in metronomes. As in the case of music-production tools, drum trainers equipped with clave-training ability would be a welcome addition to the current set of music-training tools on the market today for many percussionists and other musicians.

5.4.3 Music Teaching

Clave awareness is currently not a common part of general music education, and does not even seem to be a formal part of percussion education anywhere. The concept of clave direction would be a welcome addition to music education. Not only would it prepare musicians for playing Afro-Latin music, it would also broaden the intellectual view students and educators have of the role of rhythm in non-European musical contexts. Again, since clave consciousness is difficult to cultivate and not a regular part of many musicians' and academics' primary work, technological tools may be invaluable.

From the standpoint of pedagogy, a high-tech clave-awareness tool would allow teachers (whether in the area of percussion or not) to let students explore rhythmic possibilities without the need to come up with every pattern the student is to practice to develop a sense of clave, and also without having to provide an answer to the clave conformity of every pattern the student creates. In this way, teachers of music can incorporate this concept into their work without first having to completely master it and can even benefit from the device for their own professional development. Clave awareness can thus be incorporated into learning situations that have a different primary

focus, without requiring extensive preparation on the part of an instructor and while expanding their areas of and strategies for teaching.

5.5 Conclusion

Clave analysis is an emerging field that, when combined with other forms of machine listening, can aid musicians, listeners, researchers, instructors and students in characterizing, selecting, teaching, learning and performing a vast and increasingly popular set of genres within the Afro-Latin idiom. From the point of view of music technology, it fills a niche whose presence has been felt for some time among performers and the instrument industry.

From the perspective of multicultural music education, it focuses on a much-neglected aspect of an influential tradition and it is an excellent avenue for technology to contribute to music education, music training and their cultural diversity.

From the point of view of pedagogy, it increases the musical (rhythmic) options available to students and teachers who wish to delve into this subtle aspect of musical temporal organization. In short, clave-aware technology for music education is poised to help industry (in the form of new, exciting products), educational practices (in the form of expanding educators' opportunities and abilities) and music-making (for performers who wish to develop their clave sense), all the while contributing to the recognition of cultural diversity in the realms of both music and education.

5.6 JMTE-Suggested Citation for the Published Article

Vurkaç, M. (2011), 'Clave-direction analysis: A new arena for educational and creative applications of music technology', *Journal of Music, Technology and Education* 4: 1, pp. 27–46, doi: 10.1386/jmte.4.1.27_1

CHAPTER VI: FINDINGS/RESULTS

6.1 Results of Neural-Net Design

The results are organized so as to show the first pass through the holdout data of each parallel research track, along with any controls, followed by the results of iterations with different network seeds for the minimum number required for a small effect size.

The three-hidden-element fully connected network with default seed performed 91% correct on the holdout. The seven-hidden-element feature-reduced prestructured network based on MBR-a (using AIC followed by BIC) with default seed performed 73% on the holdout. The pure RA performance of the same model was 69% correct on the holdout.

The fit-simplified version of the feature-reduced MBR-i (Information) model, ABCQZ:ABRZ:BFJQZ:BMQRZ:CJKZ:CKRZ:FGZ:GRZ:JKMZ:JKQZ:JKRZ:JQRZ:KMZ:KQRZ, performed 86% on the holdout with the default seed. The pure RA performance for this model was 68% correct on the holdout.

The ABE-b RA search led to what turned out to be the most successful model. The ABE-b (BIC)-based model is IV:ABZ:ACZ:AJKZ:ASZ:BDZ:EFGZ:FJZ:HJZ:JKMZ:JNZ:JQZ:JRZ:PZ:QSZ:RSZ with 38 degrees of freedom.

After running “fit,” the winning ABE-i (information) model based on incremental alpha is IV:ABZ:ACZ:AGZ:AKZ:ARZ:AMZ:ASZ:BDZ:BNZ:BQZ:BRZ:CDZ:EFGZ:FJZ:FQZ:FRZ:GHZ:GPZ:GQZ:GSZ:HJZ:HPZ:HQZ:HSZ:JKRZ:JNZ:

JQZ:KMZ:MQZ: NQZ:PQSZ:RSZ. This model has 57 degrees of freedom and 32 terms corresponding to 32 hidden-layer elements.

Note that the “lousy” control network (IV:CGHZ:CKZ: ACDGHJKMNPQRSZ:CDEJMNRZ) is similar (in terms of its input-to-hidden layer connections) to a fully connected network, so we would expect it to approach the performance of a fully connected network. In fact, this model was observed to perform 77% correct on the holdout set.

6.2 Tables of Numerical Results

The initial comparison can be found in Table 8. Generalization by uninformed individuals who had access to a representative (stratified) training set ranges from 14 to 57%, with the person scoring 14% having made an assumption as to what was sought instead of building her own representative model for the data. Since this assumption was incorrect, the classification that resulted from it was very far from the target. The floor value, for lack of a better (bigger) sample, is taken to be the average of these performances, which is 36%.

The generalization of mid-level teacher/performers (professional musicians in Afro-Cuban or Afro-Brazilian music) had an even smaller sample size, since it relied on a very small group of qualified candidates. The average of this set was 61%. However, C can be considered an outlier since his specialty is Cuban and New Orleans music. The two other subjects, T and R, who specialize in Brazilian music, scored in the 70s. Nonetheless, the highest value in this small sample, 76%, is still too low to be

considered the ceiling value for neural-net performance, especially if the practical objective is a product to help musicians learn clave direction.

For the neural nets, the performance of the “lousy” control structure was indeed lousy; holdout performance dropped by 11% from the rate at which the network learned the training set. The fully connected (standard) network also lost a small amount of performance when going from learning to generalizing, though this loss was not replicated in the subsequent trials with different network seeds (Table 9).

It was not possible to find the RA-for-classification performance of the *Occam3* program in the same manner as the NNs (by initializing to new network seeds). RA generalization rates were no less than 16 percentage points lower than learning rates, and as much as 34% lower in one case.

The only hypotheses (agents, cases) that did not drop in performance when going from learning to generalization were three of the prestructured neural nets: ABE–b (BIC), ABE–i (Information), and MBR–i (Information). The latter was dropped from the final round, on account of its relatively low performance, and because affiliated RA and NN performances were weaker than those of the candidates for ABE–b-prestructured and ABE–i-prestructured neural networks.

A final prestructured network type was added to the list of candidates in order to test two ideas: the notion that RA is a fruitful way to prestructure NNs, and the question of whether *musician’s intuition* would help the problem. This latter network type is alternately called “musician’s intuition” or “MV *a priori*” because it was drawn up

based on a rhythmic analysis of clave direction prior to the start of the experiments. A portion of the results for all five network types is given in Table 9.

Table 8 summarizes the initial-pass results (for NNs, with the default seed) for training and holdout (test) for four RA-prestructured networks, four corresponding RA-only classifications, the fully connected network, the arbitrary control network (“lousy”), and the generalization performance (on a smaller, proportionally representative subset) by a group of mid-level human experts and a group of uninformed human subjects.

Table 8: Initial Results

<u>Performance on the Holdout</u>				
Fully Conn. NN	MBR-a prestr. NN	MBR-a RA	MBR-i prestr. NN	MBR-i RA
91%	73%	69%	86%	68%

Lousy (Control) NN	ABE-b prestr. NN	ABE-b RA	ABE-i prestr. NN	ABE-i RA
77%	92%	60%	92%	60%

<u>Learning Rate</u>				
Fully Conn. NN	MBR-a prestr. NN	MBR-a RA	MBR-i prestr. NN	MBR-i RA
92%	84%	85%	86%	89%

Lousy (Control) NN	ABE-b prestr. NN	ABE-b RA	ABE-i prestr. NN	ABE-i RA
86%	92%	94%	92%	94%

<u>Generalization by Mid-Level Teacher or Performer (by initial)</u>				
C	T	R	J	AVERAGE
38%	76%	70%	dropped out	61%

<u>Generalization by Uninformed Persons (by initial)</u>				
T	L	R	K	AVERAGE
35%	57%	38%	14%	36%

Table 9: A selection of Generalizing Ratio values for the five “finalist” neural-net types as exercised by different random-number seeds.

Network Type	BIC-based	Inf-based	Intuition-based	Fully Connected	Control
	1.086970	1.080152	1.068074	1.044343	0.972944
	1.087678	1.080613	1.061127	1.029993	0.975734
	1.080878	1.074258	1.056994	1.032089	0.966128
	1.089216	1.075679	1.063752	1.036391	0.963335
	1.085400	1.081697	1.068187	1.025683	0.973540
	1.076651	1.075234	1.059942	1.049741	0.994540
	1.084212	1.077006	1.060650	1.037823	0.969427
	1.083574	1.077634	1.065759	1.037884	0.971824
	1.091822	1.080004	1.063392	1.042653	0.969726
	1.088083	1.083913	1.060290	1.033939	0.972097
	1.083322	1.076618	1.059578	1.050887	0.990409
	1.088290	1.081145	1.066669	1.040359	0.967198
	1.086842	1.084380	1.066302	1.030598	0.993242
	1.080968	1.077601	1.066210	1.026675	0.954692
	1.080212	1.074848	1.058599	1.030450	0.971114
	1.078644	1.070757	1.062303	1.044119	0.979505
	1.083010	1.076998	1.061130	1.053776	0.991693
	1.084789	1.075622	1.065616	1.038267	0.977062
	1.086478	1.075885	1.060008	1.035339	0.997059
	1.082672	1.075315	1.065622	1.029511	0.970963
	1.080707	1.081925	1.061778	1.034038	0.981206
	1.083375	1.075830	1.064575	1.028405	0.977595

Network Type	BIC-based	Inf-based	Intuition-based	Fully Connected	Control
	1.081308	1.078469	1.057463	1.036695	0.969328
	1.089457	1.081782	1.066113	1.027348	0.964266
	1.083586	1.076317	1.058295	1.044980	0.981962
	1.084145	1.080303	1.064756	1.030862	0.967390
	1.081270	1.081610	1.065212	1.027638	0.970882
	1.090507	1.079083	1.064823	1.039861	0.975163
	1.089401	1.078158	1.063118	1.047287	0.984108
	1.084251	1.077309	1.065185	1.042703	0.977363
	1.076812	1.071849	1.060357	1.032840	0.972849
	1.085647	1.074380	1.064497	1.040150	0.979130
	1.080793	1.081553	1.066796	1.022347	0.962577
	1.081450	1.073957	1.058609	1.029227	0.971741
	1.087795	1.085293	1.063486	1.049581	0.984530
	1.081858	1.079377	1.065436	1.023962	0.968313
	1.079940	1.077652	1.064403	1.036792	0.978646
	1.085822	1.082358	1.058402	1.039286	0.971843
	1.084852	1.076651	1.063574	1.037346	0.973623
	1.086069	1.078446	1.059235	1.041645	0.979219
	1.085316	1.079483	1.064980	1.038978	0.975145
	1.088079	1.076734	1.061765	1.050998	0.980945
	1.084755	1.072565	1.061029	1.045296	0.991425
	1.079296	1.074037	1.059133	1.033706	0.976819
	1.084130	1.078300	1.066176	1.036382	0.971166
	1.084809	1.083307	1.062181	1.043940	0.989535

Network Type	BIC-based	Inf-based	Intuition-based	Fully Connected	Control
	1.088155	1.080077	1.065670	1.037823	0.975453
	1.095000	1.081099	1.063060	1.038151	0.985273
	1.079754	1.072143	1.062710	1.032356	0.972904
	1.084318	1.075901	1.058496	1.044031	0.993116
	1.088426	1.084313	1.061896	1.054988	0.987763
	1.083647	1.075902	1.059528	1.035090	0.970236
	1.088769	1.080498	1.062617	1.046563	0.975487
	1.083145	1.073624	1.059580	1.027071	0.970613

What follows is the ANOVA results for 10 observations from each network type. For the reader who may not be fully familiar with ANOVA, it is important to point out that “one-way ANOVA” refers not to it being a five-way comparison (which it is), but to ANOVA’s use of only one factor to make this five-way comparison. That factor is called “method,” and refers to whether BIC prestructuring, Information prestructuring, or some other method was used. No other factors are examined; for instance, neither temperature nor length of training time was considered. Similarly, “linear” refers to the ANOVA model, not to any assumptions about underlying quantities. This linear ANOVA model for “method” means is given in (46).

$$Ratio = \mu + \beta_i \times Method + \varepsilon_i \quad (46)$$

“Ratio” stands for each data point as predicted by the linear model. Residuals, reported below, distances from the predicted data point to the actual. The null

hypothesis is that $\beta = 0$ for all i , meaning all method means are the same. The calculated P value indicates whether to reject this null hypothesis or not.

Only 10 observations out of 388 suffice because, as seen in Table 9 and Figure 50, each set of the Generalizing Ratio data has very low variance. Any findings of statistical significance for sets of 10 are strengthened by including all observations as given in Section 6.4, even though this runs the risk of bloating the analysis for no statistical gain: “The study should be made as simple as possible, but no simpler” [132, p. 1456].

Discarding collected data is worse than bloating the study. The reason for the excess of data is that different guidelines were consulted for the statistical power. Cohen’s recommended number for seeking a small (or larger) effect in a five-way comparison was $N = 393$ [156]. However, as has been discussed in the Statistics literature, this is a recommended N for studies of a sociological, psychological, or biological type where there is a lot of variation among samples (behaviors, biochemistry, etc.). Much smaller numbers (on the order of 10) are recommended for computational studies.

Given a choice between 8 and 393 random-number-seed repetitions, the researcher chose to do 393⁹⁴ because, if the greater number turned out to be necessary after only doing 8 or 10, it would be difficult *and* scientifically dishonest (in terms of statistical significance) to go back and conduct hundreds more experiments. If, on the

⁹⁴ but could only save the results for between 385 and 388 because disk space ran out.

other hand, the smaller number turned out to be sufficient (as it did), extra data could not hurt from the point of view of knowledge of the networks' performance (and indeed provided insight into both NN and RA performance, as discussed in the conclusion). The only disadvantage in having this extra data, from the statistical point of view, is that it may appear to be "significance-chasing." We can dismiss this potential concern for two reasons: 1) Statistical significance *was* found at the target level even with $N = 10$; and 2) The excessive data were collected as part of one experimental setup, without ulterior motives, and as established by Cohen [156].

Consequently, we discuss the statistical analysis of the $N = 10$ set in Section 6.3, in order to demonstrate that results were already statistically significant at that point. To provide full scientific disclosure, the analysis of all the collected data is discussed in Section 6.4.

6.3 Preliminary Statistical Analysis

The analysis below shows Dunnett and Bonferroni adjustments for comparing multiple means, as well as the standard deviations and confidence intervals for the data. First, Table 10 provides the list of Generalizing Ratios (GRs) for ten random-number seeds for the five network types. The necessary assumptions for ANOVA (independence and equal variances) are discussed below.

Table 10: Generalizing Ratios used in the initial ANOVA, with arbitrary identifiers.

Method	Arbitrary Observation Identifier	Generalizing Ratio
BIC (BIC-based ABE RA prestructuring)	1	1.08697
BIC (BIC-based ABE RA prestructuring)	2	1.087678
BIC (BIC-based ABE RA prestructuring)	3	1.080878
BIC (BIC-based ABE RA prestructuring)	4	1.089216
BIC (BIC-based ABE RA prestructuring)	5	1.0854
BIC (BIC-based ABE RA prestructuring)	6	1.076651
BIC (BIC-based ABE RA prestructuring)	7	1.084212
BIC (BIC-based ABE RA prestructuring)	8	1.083574
BIC (BIC-based ABE RA prestructuring)	9	1.091822
BIC (BIC-based ABE RA prestructuring)	10	1.088083
INF (Information-based ABE RA prestructuring)	1	1.080152
INF (Information-based ABE RA prestructuring)	2	1.080613
INF (Information-based ABE RA prestructuring)	3	1.074258
INF (Information-based ABE RA prestructuring)	4	1.075679
INF (Information-based ABE RA prestructuring)	5	1.081697
INF (Information-based ABE RA prestructuring)	6	1.075234
INF (Information-based ABE RA prestructuring)	7	1.077006
INF (Information-based ABE RA prestructuring)	8	1.077634
INF (Information-based ABE RA prestructuring)	9	1.080004
INF (Information-based ABE RA prestructuring)	10	1.083913
MV (Musical intuition-based prestructuring)	1	1.068074
MV (Musical intuition-based prestructuring)	2	1.061127
MV (Musical intuition-based prestructuring)	3	1.056994
MV (Musical intuition-based prestructuring)	4	1.063752
MV (Musical intuition-based prestructuring)	5	1.068187
MV (Musical intuition-based prestructuring)	6	1.059942
MV (Musical intuition-based prestructuring)	7	1.06065
MV (Musical intuition-based prestructuring)	8	1.065759
MV (Musical intuition-based prestructuring)	9	1.063392
MV (Musical intuition-based prestructuring)	10	1.06029
Full (Fully Connected NN)	1	1.044343
Full (Fully Connected NN)	2	1.029993
Full (Fully Connected NN)	3	1.032089
Full (Fully Connected NN)	4	1.036391
Full (Fully Connected NN)	5	1.025683

Full (Fully Connected NN)	6	1.049741
Full (Fully Connected NN)	7	1.037823
Full (Fully Connected NN)	8	1.037884
Full (Fully Connected NN)	9	1.042653
Full (Fully Connected NN)	10	1.033939
Lousy (Arbitrarily Structured Control Network	1	0.972944
Lousy (Arbitrarily Structured Control Network	2	0.975734
Lousy (Arbitrarily Structured Control Network	3	0.966128
Lousy (Arbitrarily Structured Control Network	4	0.963335
Lousy (Arbitrarily Structured Control Network	5	0.97354
Lousy (Arbitrarily Structured Control Network	6	0.99454
Lousy (Arbitrarily Structured Control Network	7	0.969427
Lousy (Arbitrarily Structured Control Network	8	0.971824
Lousy (Arbitrarily Structured Control Network	9	0.969726
Lousy (Arbitrarily Structured Control Network	10	0.972097

The first test is a one-way ANOVA (“ratio versus method”) to see if any of the four means is statistically significantly different from that of the reference model. Since this gave us the go-ahead, we use the Dunnett ANOVA to compare each mean separately against the reference mean. Table 11 shows a high degree of statistical significance (in fact, the maximum possible) and the associated R^2 values indicate that most of the variability in the data set is explained by the linear ANOVA model.

Table 11: One-way ANOVA: Ratio versus Method

Source	Degrees of Freedom	Sum-Squared	Mean-Squared	F	P
Method	4	0.0831288	0.0207822	631.08	0.000
Error	45	0.0014819	0.0000329		
Total	49	0.0846107			

Figure 50 gives all but one of the essential pieces of information about the trials. The means and standard deviations of the GRs for each network type are listed on the

left of the figure. The BIC-based prestructured network (“BIC”) has the highest mean GR, followed closely by the other RA-prestructured network. As would be expected from browsing the data in Table 9, the intuitively prestructured network (“MV”) also outperformed the fully connected network in terms of its mean GR, as well as performing more consistently (having half as much spread), as indicated by the standard deviation. (This is discussed further below). The 95% confidence intervals (CIs)—which give a range within which the true mean is expected to lie with 95% likelihood—are tightest for BIC and MV, so we are highly confident that the true (population) mean is close to the observed (sample) mean.

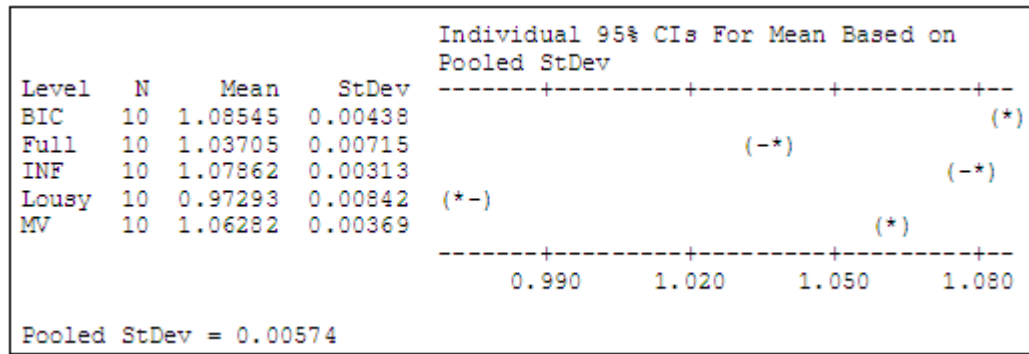


Figure 50: 95% Confidence Intervals for One-Way ANOVA, individual comparisons of Prestructured Network Types against Fully Connected.

The target analysis for the present research, the five-way Dunnett ANOVA to compare four prestructured network types against the standard model (fully connected MLP) reveals that the mean values of the GRs are each different at a statistically significant level from that of the reference (fully connected) network (cf. Table 12).

Table 12: Grouping Information Using the Dunnett Method of ANOVA

Level	N	Mean	Grouping
Full (control)	10	1.037054	A
BIC	10	1.085448	
INF	10	1.078619	
MV	10	1.062817	
Lousy	10	0.972930	
Means not labeled with letter A are significantly different from control level mean.			

Dunnett's comparisons with a control

- Family error rate = 0.05
- Individual error rate = 0.0149
- Critical value = 2.53
- Control = level (Full) of Method

Table 13 and Figure 51 show the 95% confidence intervals for the distance from each prestructured-net mean to the reference mean. (We expect the true difference to lie within this interval.) Hsu's comparison seeks to identify the best alternative with a target rate of 0.05. This clearly indicates that "BIC" was the best network type.

Table 13: Confidence Intervals for Treatment Mean minus Control Mean

Level	Lower	Center	Upper
BIC	0.041898	0.048394	0.054891
INF	0.035069	0.041565	0.048062
Lousy	-0.070621	-0.064124	-0.057628
MV	0.019266	0.025763	0.032259

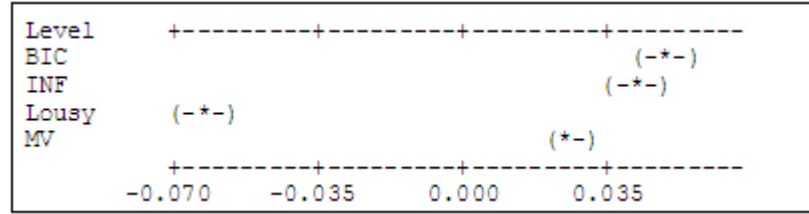


Figure 51: Confidence Intervals for Treatment Mean minus Control Mean

Hsu's MCB (Multiple Comparisons with the Best)

Family error rate = 0.05

Critical value = 2.22

Figure 52 and Figure 53 justify the use of ANOVA (for the most part). The residuals show deviation from the normal (Gaussian) distribution, which is an underlying assumption for ANOVA. The random sampling of the design and holdout sets for the NN experiments and the large sample sizes used do not guarantee a Gaussian distribution at this stage because the population under test is generalization performances, not pattern vectors. Even if the large number of experiments could guarantee normality through the central-limit theorem, because the GRs are ratios, we would at best expect an F or a χ^2 distribution. Thus, Bartlett's test for equal variances (below) is not strictly applicable, but its inapplicability (thus the inappropriateness of doing ANOVA) is tempered by the fact that the two ways in which assumptions are not met are both so in a rationally insignificant way: 1) The deviation from normality is very small, and 2) it is due to a single outlier among the “lousy” set (not even an important part of the experimental setup).

Furthermore, Levene's test for the same purpose is applicable, and the P value of 0.444 indicates a high degree of confidence that any observed differences in variance are due to sampling error. This, in turn, is the justification needed for running any ANOVA. (See recommendations in Appendix D.)

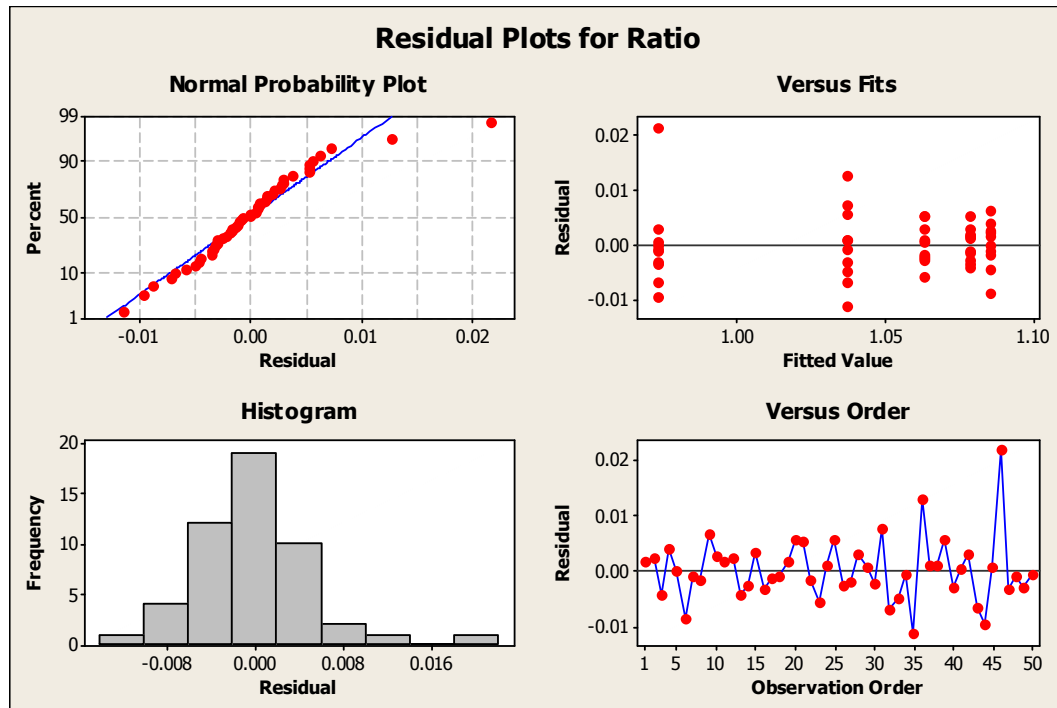


Figure 52: Residual Plots for Ratio

ANOVA presupposes a normal distribution of residuals, and that cases (i.e., the GRs of the network types) are independent.

The independence condition deserves attention: Are the cases probabilistically dependent because the networks are initiated with the same random-number seeds and

are trained and tested on the same data (in random order), or are they probabilistically independent because the elements that play the largest role in determining generalization performance (the hidden-layer elements) are in unrelated configurations?

We assume the latter because the network structure is the most important determining factor as long as the training and test data are judged to be uniformly representative of the environment, which they are.

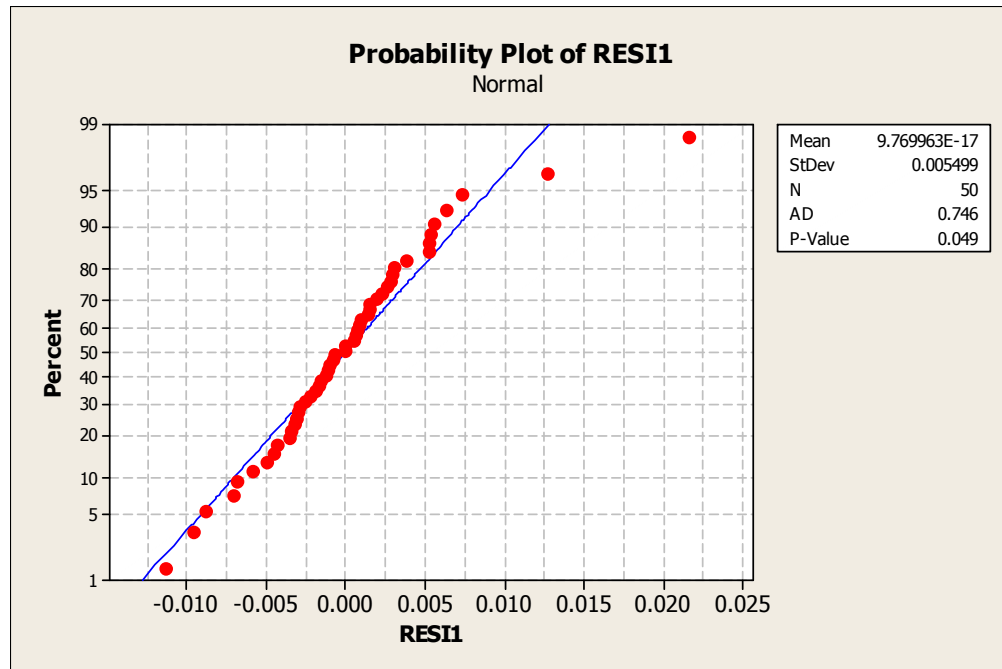


Figure 53: Test of Normality of Distribution

Bartlett's Test for Equal Variances (for a normal distribution)

Test statistic = 12.30, P -value = 0.015

Levene's Test for Equal Variances (for any continuous distribution)

Test statistic = 0.95, P -value = 0.444

The confidence intervals for standard deviations (Table 14 and Figure 54) support the Levene's Test result that variances are likely to be equal. This can be seen by placing a vertical line at about 0.0051 in Figure 54.

Table 14: 95% Bonferroni Confidence Intervals for Standard Deviations

Method	N	Lower	StDev	Upper
BIC	10	0.0027032	0.0043763	0.0099676
Full	10	0.0044189	0.0071541	0.0162942
INF	10	0.0019362	0.0031346	0.0071394
Lousy	10	0.0052007	0.0084197	0.0191769
MV	10	0.0022782	0.0036884	0.0084007

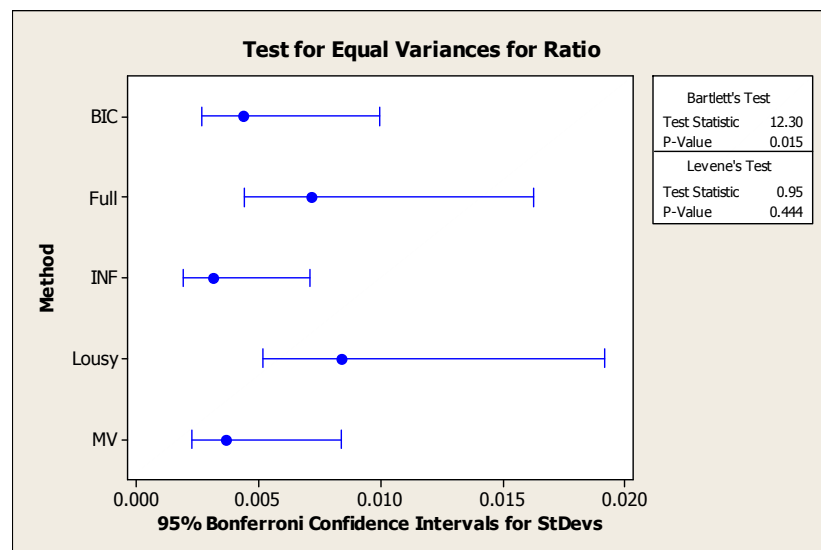


Figure 54: 95% Bonferroni Confidence Intervals for Standard Deviations.

6.4 Full Statistical Analysis

The same set of analyses has been performed for the complete data set. Again, we have a “one-way” ANOVA, meaning that the five-way comparison is made on only one factor, which is the method used to obtain the data, not how long it took, or under what physical or other circumstances the data were acquired.

Again, the Dunnett and Bonferroni adjustments for comparing multiple means are employed, and confidence intervals are obtained for both means and standard deviations. Further analysis is done to investigate whether the ANOVA assumptions hold up.

The P value resulting from the initial test is arbitrarily close to 0, and the linear ANOVA model is able to explain 96% of the variation in the data (Figure 55). These findings provide confidence about running the ANOVA analysis on the data at hand.

Source	DF	SS	MS	F	P
Method	4	3.075616	0.768904	11351.47	0.000
Error	1930	0.130731	0.000068		
Total	1934	3.206347			
R-Sq = 95.92% R-Sq(adj) = 95.91%					

Figure 55: Ratio versus Method for Full Analysis.

Next, we see the means and standard deviations of the Generalizing Ratios (GRs) for each network type (Figure 56).

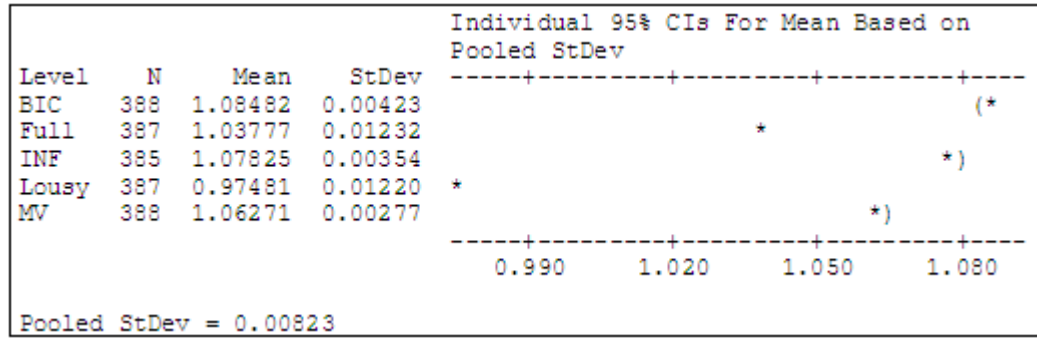


Figure 56: Means, Standard Deviations, and 95% Confidence Intervals for the Full Analysis.

Just as in the $N = 10$ case, we see that the BIC-based prestructured network (“BIC”) has the highest mean GR, followed closely by the other RA-prestructured network (Information). The network prestructured according to musician’s intuition (“MV”) is also a high performer with low standard deviation, and indeed the overall raking is exactly the same as in the $N = 10$ analysis above (Section 6.3).

The 95% confidence intervals (CIs)—which give a range within which the true mean is expected to lie with 95% likelihood—are again tightest for BIC and MV. Even with the other three network types, confidence intervals are so tight that the display does not even have room for two dashes around the mean. We can, therefore, be highly confident that the true (population) mean is close to the observed (sample) mean for each case.

Five-way Dunnett ANOVA for comparing four prestructured network types against the standard model (Figure 57), and the five-way Bonferroni adjustment (Figure

58) both reveal that the mean values of the GRs are each different at a statistically significant level from that of the reference (fully connected) network.

Grouping Information Using Dunnett Method and 95.0% Confidence			
Method	N	Mean	Grouping
Full (control)	387	1.0	A
BIC	388	1.1	
INF	385	1.1	
MV	388	1.1	
Lousy	387	1.0	
Means not labeled with letter A are significantly different from control level mean.			

Figure 57: Dunnett ANOVA showing that all four network types are statistically significantly different from fully connected.

Grouping Information Using Bonferroni Method and 95.0% Confidence			
Method	N	Mean	Grouping
Full (control)	387	1.0	A
BIC	388	1.1	
INF	385	1.1	
MV	388	1.1	
Lousy	387	1.0	
Means not labeled with letter A are significantly different from control-level mean.			

Figure 58: Bonferroni Adjustment showing that all four network types are statistically significantly different from fully connected.

The 95% confidence intervals and *P* values for Dunnett (Figure 59) and Bonferroni (Figure 60) follow.

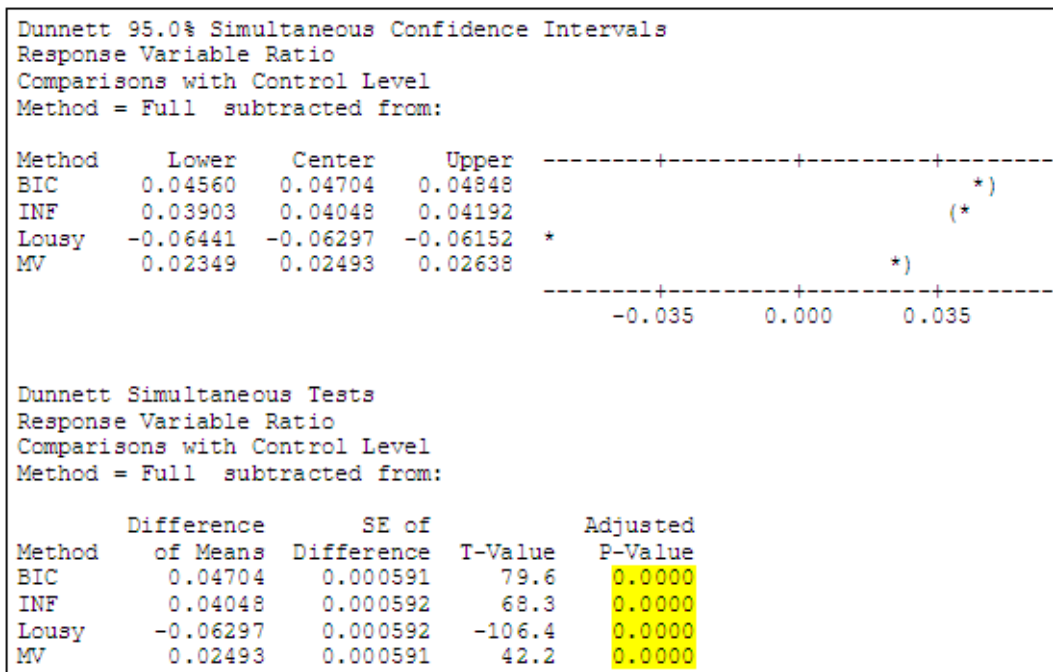


Figure 59: Dunnett 95% Confidence Intervals and *P*Values.

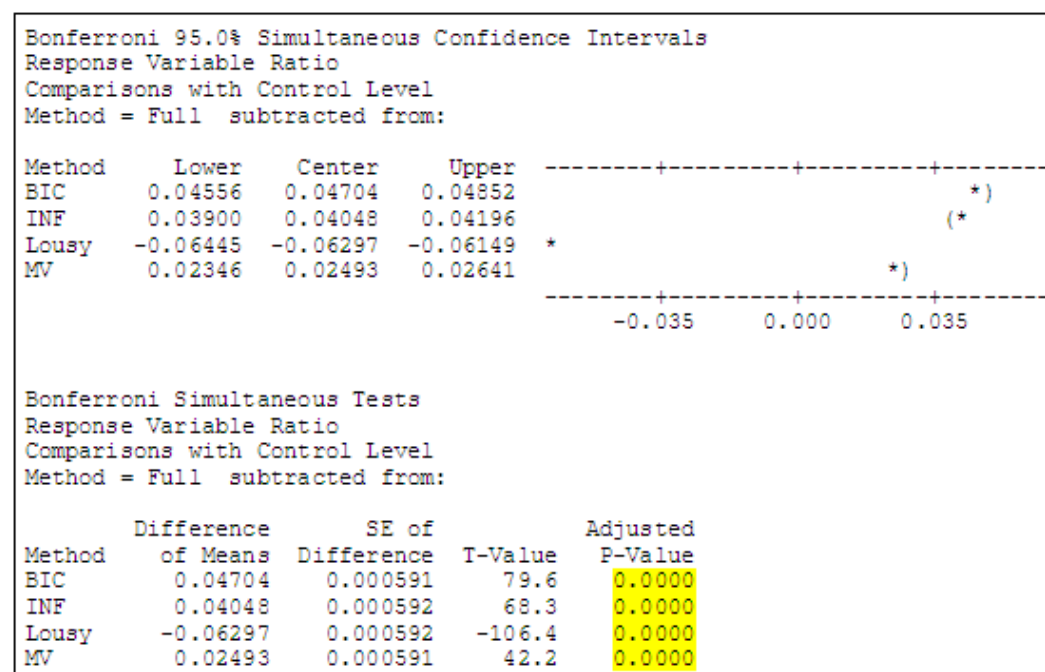


Figure 60: Bonferroni 95% Confidence Intervals and *P*Values.

The confidence intervals have only gotten tighter, such that only one parenthesis fits in the required space, and no room for dashes around the means.

In addition to these analyses, the Tukey method was used so as to have an all-way comparison. The pairwise confidence intervals for Tukey appear the same as the previous analyses, so, to avoid redundancy, they are not displayed. However, the Tukey grouping information for means is not redundant, so it is given in Figure 61.

Grouping Information Using Tukey Method and 95.0% Confidence		
Method	N	Grouping
BIC	388	A
INF	385	B
MV	388	C
Full	387	D
Lousy	387	E
Means that do not share a letter are significantly different.		

Figure 61: Tukey pairwise comparisons show statistically significant differences among all means. (All Tukey *P* values are arbitrarily close to 0.)

As in the preliminary analysis, the assumptions for being able to perform the ANOVA must be checked. Both Bartlett's and Levene's tests result in *P* values that are arbitrarily close to 0. However, there are now four outliers (out of 1935 data points compared to one out of ten in the preliminary analysis) as shown in Figures 47 and 48.

The outliers are the following:

- Lousy: 0.0809 (very lousy!)
- Full: 0.931 (degradation instead of improvement)
- Full 0.934 (degradation instead of improvement)
- Full: 1.120 (best overall performance!)

With the residuals in Figures 47 and 48 showing such a close match with normality for 1931 out of 1935 data points, even though the statistical calculations suggest otherwise, rational examination of the data implies that these few outliers (especially since none are in the prestructured networks' performances) should not invalidate the entire study and its analyses. It is, however, poignant to inspect the outliers themselves.

The “lousy” outlier has the worst generalization performance of all 1935 networks (1940, if the default five are included).

The next two outliers are not enough to hurt the conclusion of the research, but in a small way, may actually bolster it: The only networks that show any degradation of generalization performance are of the “lousy” and the fully connected classes. While the former was expected, or at least hoped for⁹⁵, the latter is another piece of support for the prestructuring hypothesis in that it is possible for the fully connected networks to *overfit* even after k -fold cross-validation design.

The fourth and final outlier is the highest performing network in the entire study, and this calls for consideration. Could it be that the notion of prestructuring should be discarded, and this high-performing (fully connected) network be selected in its place?

The argument for the answer “no” has two parts.

⁹⁵ One would hope that the “lousy” networks, which were randomly structured, would under-perform all other network types. Otherwise, the effort of information-theoretically prestructuring (or indeed, prestructuring in any informed fashion) would be worthless by virtue of performing worse than a random structure.

1. The number “1 out of 1935” is 0.0005168, and the number “1 out of 387” is 0.002584, both of which are well within the statistical-significance target of 0.05, and such an occurrence may be interpreted as a chance occurrence due to sampling.

2. The training and test data together make up roughly 16% of the I/O space for the present problem, and the holdout data makes up about a fifth of that. This means there are roughly 54,700 I/O pairs yet to be tested (or used for training, but not both). It is likely that this fully connected network’s performance is, again, due to sampling error⁹⁶, and could change under different training and testing conditions.

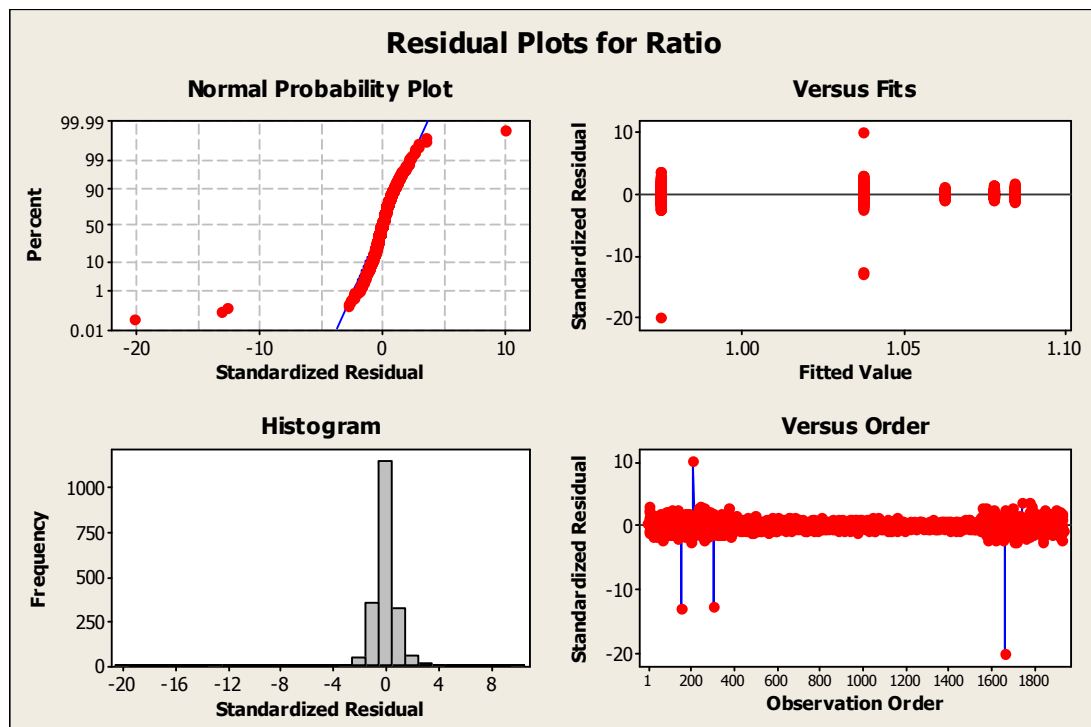


Figure 62: Results of the tests of normality for the full analysis.

⁹⁶ Sampling error is not a matter of blame, but a systemic necessity of sampling a population.

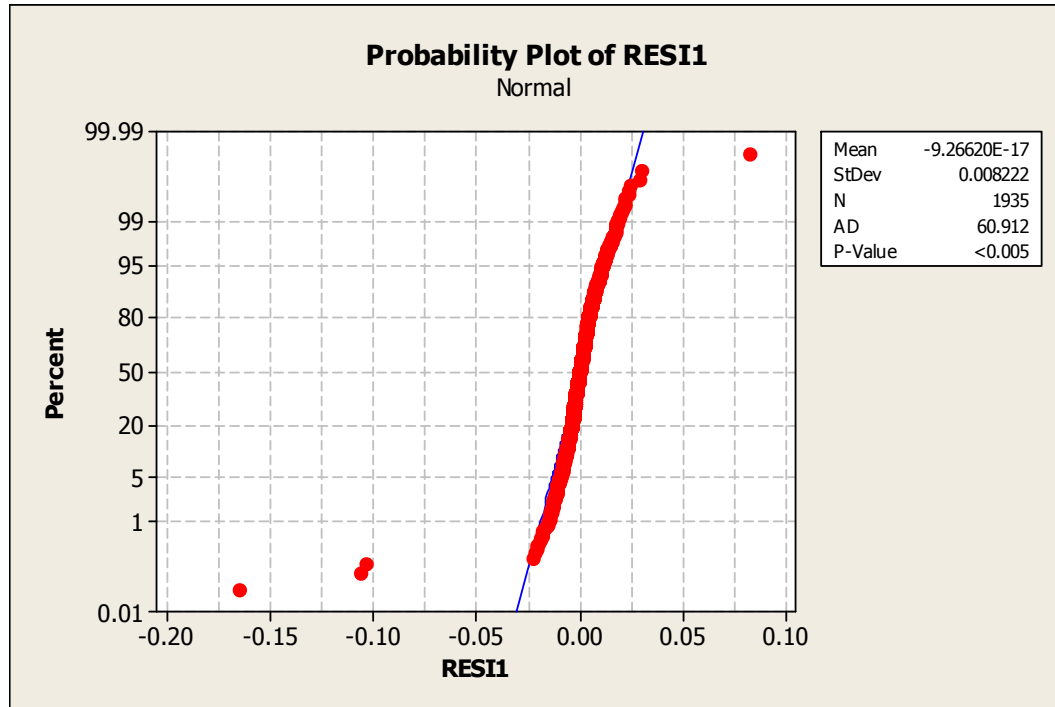


Figure 63: The four outliers (out of 1935 data points) that lead to doubts on accepting the normality assumption for ANOVA.

Table 15: 95% Bonferroni Confidence Intervals for Standard Deviations for the Full Analysis

Method	N	Lower	StDev	Upper
BIC	388	0.0038680	0.0042281	0.0046563
Full	387	0.0112690	0.0123195	0.0135689
INF	385	0.0032411	0.0035440	0.0039045
Lousy	387	0.0111585	0.0121987	0.0134359
MV	388	0.0025323	0.0027681	0.0030484

In Table 15 we see that the confidence intervals for the prestructured networks have all tightened up, and the confidence intervals for the fully connected and “lousy” widened in comparison with the preliminary analysis (Table 14). This can be observed visually in Figure 64.

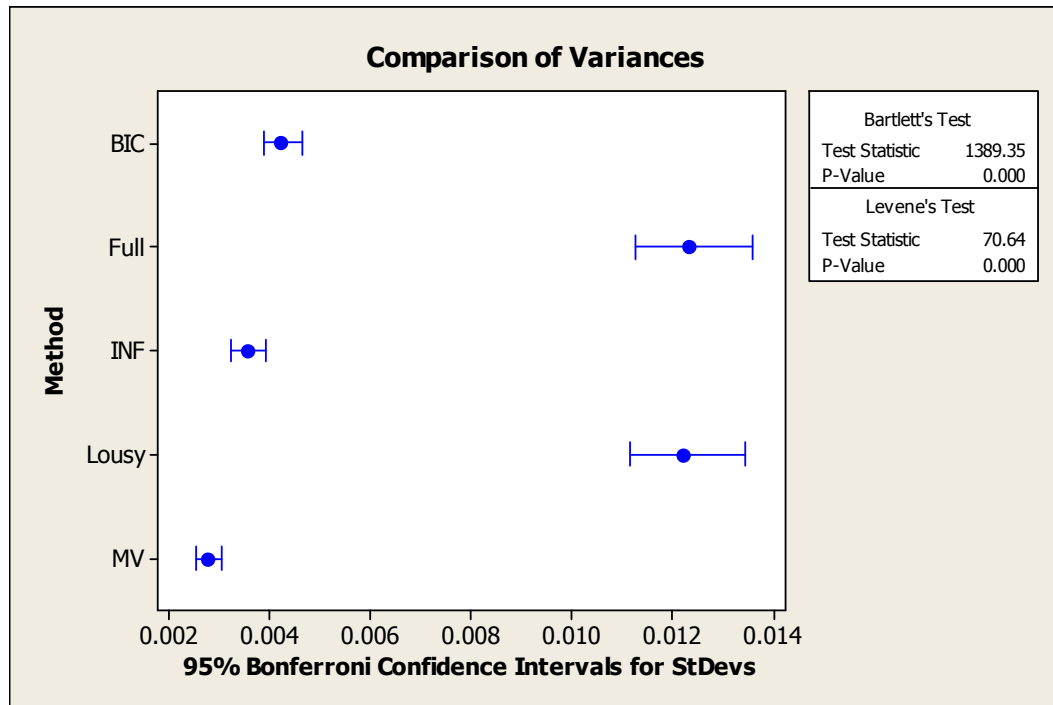


Figure 64: Comparison of Variances for the Full Analysis.

Since this difference in standard deviations implies a disparity between the assumptions of Dunnett ANOVA and the affiliated adjustments and comparisons, a form of statistical analysis is needed that is resilient to variance heterogeneity. There are two choices, both seldom used: the Bayesian Waller-Duncan test which balances statistical power and statistical significance, but is typically reserved for cases with small N , and Welch's ANOVA.

For Welch's ANOVA, the P value and the percentage of variance in the data explained by the method (in this case given by W^2 , equivalent to R^2) are the same as in the previous analyses. Regardless of variance heterogeneity, we are able to conclude that at least one of the group means differs from the overall mean. As above, Tukey's pair-

wise comparisons can be employed for a detailed analysis. The results once again indicate that each of the individual methods fall into their own category; i.e., their confidence intervals do not overlap. Given the Welch's ANOVA and Tukey results that BIC-prestructuring outperforms the other methods statistically significantly, we employ Hsu's comparison with the best one last time to identify the method or methods that have the highest GR. The results of this test demonstrate that the best method is the BIC-prestructured network with no other methods sharing this "best" category. It should be noted that under the Bonferroni, Sidak, and Dunnett's multiple comparisons, the classification of the results turned out the same, suggesting that BIC is the optimal method.

Table 16: Welch's ANOVA, as compared with Dunnett ANOVA.

One-way ANOVA: Ratio versus Method					
Source	DF	SS	MS	F	P
Method	4	3.075616	0.768904	11351.47	<0.001
Error	1930	0.130731	0.000068		
Total	1934	3.206347			
Measures	S = .00823	R ² = 95.92%		R ² (adj) = 95.91%	
Welch's ANOVA					
Numerator DF		4		P <0.001	
Denominator DF		924.1926612			
F statistic		8884.234954			
W ²		95.91%			

6.5 Discussion of Findings

We have found, to the best of our ability to infer, evidence to support the following claims:

- Prestructured neural networks have better generalization performance than fully connected neural networks, and that this claim is statistically sound.
- While statistically sound, the degree of generalization improvement from fully connected (standard) to the best-performing prestructured network is small—on the order of a 5% improvement. Whether this is practically significant depends on the application. For musical applications, the expected improvement may or may not be worth the investment in development time, but for neural networks employed in circumstances where lives are at stake, the family of any additional survivor due to that 5% improvement will likely find it worthwhile.
- Successful prestructuring of neural networks is feasible with Reconstructability Analysis (RA).
- RA-prestructured networks outperform not only fully connected neural nets, but also Reconstructability Analysis itself (when the latter is used for classification).
- There is structure in the clave-direction concept which can be extracted and understood.
- Under the right preprocessing conditions, clave direction can be learned by artificial neural networks to a higher degree of reliability than by human experts.

In support of these claims, we list again the means of the Generalizing Ratios (the holdout percentage divided by the learning percentage, adjusted for training- and test-set sizes) for the five network types:

- BIC-prestructured: 1.085
- Information-prestructured: 1.078
- Intuition-prestructured: 1.063
- Fully connected: 1.038
- Arbitrarily prestructured (“Lousy”): 0.975

Statistical significance was found at the 95% level such that BIC-prestructured networks performed statistically significantly better than each of the others, separately, by Welch, Dunnett, Bonferroni, Tukey, and Hsu’s tests. The information-prestructured networks performed statistically significantly better than fully connected and arbitrarily structured networks, according to Tukey’s all-way comparisons. By the same test, the intuition-prestructured networks did statistically significantly better than fully connected and arbitrarily structured, and the fully connected (standard!) only outperformed the arbitrarily structured networks (except for one outlier).

The expression “statistically significantly” is used repeatedly in the above interpretation of the statistical analyses because statistical significance is not necessarily practical significance, as discussed at various points in this dissertation. The question remains whether these findings are of practical significance to the Neural Networks field, and also whether the effect sizes are small or not.

For the former, it may be argued that a mean GR improvement of five percentage points may not be significant, and may not justify the investment in Reconstructability Analysis. Whether this is the case or not depends on the field of application. For an entertainment application, for instance, the generalization improvements shown may not be of importance. However, the applicability of neural networks typically includes many fields in which lives are at stake. In such uses, a five-percentage-point improvement could mean saving a life or two, and *that* is significant⁹⁷.

The latter concern concerned effect sizes, and again we have practical effect size and statistical effect size. In this case, the tables are turned, and the effect sizes observed are practically small, but statistically not so: Statistical effect size has to do with the ratio of a difference of means and the standard deviation. Hence, the effect sizes of small (0.2), medium (0.5) and large (0.8) are multipliers of the standard deviation. The pooled standard deviation for the present study is 0.008 (cf. Figure 56). A large effect, by this measure would be a difference of 0.0064. ***Since all of the differences we have found among means are an order of magnitude greater than this, the “pooled” interpretation of statistical effect sizes is statistically very large.***

For a worst-case scenario, we take the upper end of the 95% confidence interval for the largest standard deviation in the study: 0.0135689 for fully connected. For a large effect size, this leads us to seek a difference of means of 0.01085512. The smallest

⁹⁷ Although the topic is quite serious, one may recall the statistician joke: If the heads of ten mice were cut off, and one survived for months afterwards, what would a statistician call this result? The answer is “not significant.”

difference of means is between BIC and Information, and is equal to 0.00657. This number is very close to the *worst-case* (fully connected) medium effect size of 0.00678. Since it *is* smaller, however, we must report that we have found a *small* effect size in the difference in performance between BIC and Information, at least by this worst-case approach of using the fully connected networks' value for all differences.

By similar analyses, in the worst-case scenario, we find medium effect sizes between each of the three prestructured networks and the industry standard (fully connected), and a large effect size from each of the three prestructured networks to the “lousy” control, but only a medium effect size between the fully connected and the ‘lousy,’ further reinforcing the conclusion that prestructuring is preferable (from the point of view of generalization performance) to fully connected architectures.

Furthermore, a more fair approach to the effect-size analysis (using the worse of the two standard deviations for any given pair currently undergoing the comparison, for example) is certain to make some of these small effect sizes come up medium or large.

The question remains whether these effect sizes defined as “medium” or “large” in Statistics are actually medium or large to domain-expert interpretation. Taking a little Bayesian spin at this point, these differences in means are not as large as were expected by the researcher prior to carrying out the experiments and working out the numbers. The promise of Occam’s Razor, the Lendaris–Stanley conjecture, and the arguments from probability (both the present author’s and MacKay’s [148, p. 15]) led to an

expectation of more drastic improvements. The lack of such drastic differences in performance is a testament to the multilayer perceptron's capability to learn complex problem domains. Consider that:

1. There is very small variance in learning and holdout (test) performances within any one network's group of runs, and similarly small (though not as small) variance in the entire study;
2. Even the truly randomly structured "lousy" network achieved some degree of learning and generalization of this particular interpretation of clave direction that was better than informed human agents working with smaller but stratified and representative training and test sets (although they clearly overtrained and overfit).

Statement 1 suggests that MLPs converge to very similar solutions, which in this case have also been shown to be better solutions than reached by mid-level human experts. (Optimality was not targeted, only practical usefulness.) The MLP paradigm is seen to be invariant to starting conditions.

Statement 2 suggests that MLPs can reach a reasonably good solution even when not provided with the proper resources (although to get there, the "lousy" networks consistently, without exception, overfit to the training data). The "lousy" network can be thought of as being analogous to a properly designed network (perhaps fully connected) that suffered damage and lost a large portion of its connections. Just as we know that wet neural networks (biological brains) can reallocate resources to crucial

tasks, the intentionally “damaged” networks in the present study may have shown similar resilience in attempting to solve the clave-direction problem.

Hence, in addition to the evidence of prestructuring success (in terms of generalization improvement), we have also found incremental support for the MLP paradigm as a successful pattern-recognition methodology, even in the absence of informed prestructuring.

CHAPTER VII: CONCLUSION

7.1 Prestructuring Performance

Prestructuring has been shown to improve generalization performance (as measured by the Generalizing Ratio in neural networks of the MLP (multi-layer perceptron) type.

Specifically, networks prestructured through 1) Reconstructability Analysis using BIC as the model-selection criterion, 2) Reconstructability Analysis using Information as the model-selection criterion, and 3) musical intuition based on knowledge of the *partido-alto* paradigm⁹⁸ all outperformed the standard approach (fully connected neural network optimized under the conditions stated in Chapter I) *and* the arbitrarily created “lousy” control. Furthermore, they did so with statistically significant results at the 5% level, tight confidence intervals, and mostly medium-to-large effect sizes.

Moreover, the two RA-based networks outperformed the musical-intuition-based network, and the BIC-based RA network outperformed all other networks again with statistically significant results.

⁹⁸ limited to a form that leads to an apples-to-apples comparison with the RA-based candidates.

7.2 Statistical Significance

Despite all the problems with statistical significance, including the inherent ones, hypothesis testing is nonetheless a necessary procedure⁹⁹. A few real-life examples from a report on school-age delinquency [171] will illustrate the point in terms of what can go wrong when statistical significance is not pursued at all:

- Reported result: 25% of the 16- and 17-year-olds in the Portland, Maine, Bayside East Housing Project were out of school.
Actual data: *Only eight children were surveyed; two were found to be out of school.*
- Reported result: Of all the secondary[-]school students who had been suspended more than once in census tract22 in Columbia, South Carolina, 33% had been suspended two times and 67% had been suspended three or more times.
Actual data: *CDF found only three children in that entire census tract who had been suspended; one child was suspended twice and the other two children, three or more times.*

It is clear from such misuse that statistical significance is a valid pursuit. What has been criticized in the literature (on Statistics, Statistical Learning, Computer Science, Education, and Neural Networks) is the pursuit of statistical significance as the sole criterion in determining the meaning *and* relevance of research results without due consideration for the researchers' or analysts' background knowledge, insight, common sense, and rationality.

The sole pursuit of statistical significance becomes problematic at the point where a large enough collection of data delivers a statistically significant result no matter

⁹⁹ The aforementioned problems, discussed throughout this document (especially Chapter IV and Appendix C), make it so that statistical significance is a necessary but not sufficient measure of scientific validity.

how meaningless the data are, how faulty their sampling was, or how irrelevant the particular collection is to the research question.

The present results that neural nets based on partial and unbalanced representation of input relationships can learn clave direction (in the particular sense employed here) and generalize more successfully than informed, trained human practitioners suggests one or both of two possibilities:

- that multilayer perceptrons have once again been shown to be remarkable in their ability to learn, represent and generalize categorical notions in a domain of human endeavor, and
- that certain rhythmic figures (clave schemata and their placement along the cycle) are more strongly indicative of clave direction than others, so as to not require the entire schema to be experienced (similar to recognizing an object or idea when seeing only part of an image or word).

7.3 Implications for the Use of Reconstructability Analysis in the NN Field

The present research suggests that prestructuring based on Reconstructability Analysis is effective not only for the constructed examples covered in the literature review above, but also for the real-world case of noisy data with a cultural component.

The improvement seen in the prestructured networks may be considered incremental, but the presence of the improvement is undeniable. It is reasonable to say that a greater degree of generalization improvement was expected at the outset of the

study. This may yet be achieved with further studies that take more parameters into account or that involve additional model-search, model-selection or prestructuring techniques. *Nonetheless, the two goals set for the present research: showing generalization improvement as a result of prestructuring, and showing this at a statistically significant level for a preset significance target, have both been achieved.*

7.4 General Implications for Neural Networks

The low variance in the results obtained by networks of each type as trained from different starting points given by the 388 random-number seeds indicates that the MLP paradigm is highly resilient to random starting points for network training. Whether prestructured or not, networks trained from different starting points (seeds) maintain a uniform generalization performance. This is further evidence for the stability of the MLP paradigm (the connection scheme: hidden layer; the learning algorithm: backpropagation; and the element type: sigmoid).

7.5 Musical and Music-Technological Insight

The better performance of the BIC and Information models over the first-level musical-intuition model supports the author's claim (in this and other works) that clave direction is a nonlinear process that involves more than the recognition of certain schemata or the comparison of the presence of certain key onsets, but features an unpredictable interaction between such schemata and such onsets as they combine in all available forms. A deeper level of musical insight might be available, perhaps by

incorporating more hidden layers, but this would require similar structural changes to the prestructuring process for all other network types used in the study and is not part of the scope of the present research. Nonetheless, the apparent conclusion is that musical intuition based on an intellectual understanding of clave direction (as manifested in the *partido-alto*-based *carioca* tradition) was insufficient to express even the degree of intricacy captured by the BIC and Information models through Reconstructability Analysis. This also serves as rigorous computational evidence in favor of the claim of intricate complexity for the clave concept (cf. Appendix A).

7.6 Recommendations Concerning the Methodology of Future Studies and Focus of Future Development

What follows is a list of three suggested approaches for furthering the present study, followed by a design idea for a new neural-net paradigm inspired by some of the incidental findings of the present research.

For further developing prestructuring-for-MLPs, taking membership degrees into account in the encoding of output categories and the selection of data would be the first additional step. This requires basing such decisions on a solid foundation of behavioral (cognitive-psychological) and engineering (pattern recognition) studies involving a broader literature review.

Similarly, one could look into the results of the same approach with other output categories than the one used, and the other holdout option than the one used in the present study for verification or possible falsification of the present results.

The study could also be performed on data from a very different realm of inquiry.

As for moving in a different direction, that of design, the present study suggests developing an Adaptive Resonance Theory-inspired Fuzzy Neural Network that takes membership degrees into account in the same way a human expert would.

This idea was borne out of the discovery of the membership degrees and subsequent deliberation over how they are used in the human classification of patterns for their clave direction. It may prove useful because other problems in the field of Computational Intelligence may involve qualitative gradations in the relevance and salience of available data. Such a development requires further investigating the process of classification under such conditions as carried out by reliable experts in various fields, and building the algorithm around the findings. Hence, the key feature of such an NN paradigm would be accessibility for other areas of application in which distinctions between classes are not (always) cut-and-dried. And although this is a type of the currently popular hybrid of Fuzzy Neural Networks, this paradigm would incorporate Fuzzy notions and the plasticity of ART without being a Fuzzy-ARTMAP. In a successful implementation, the final classifier (hypothesis, agent, or network) would give the greatest weight to those rules, models, and connections that have to do with the strong and very-strong patterns. However, the gestalt nature of clave means that even the weakest patterns have to be considered, as do the least significant variables. The design question would be how to develop a system that takes some interactions

(models, variables, connections) more seriously than others, but manages to incorporate the effect of the less important ones as well.

7.7 Limitations

Although it was anticipated that computing power would no longer be an obstacle to prestructuring research with large noisy data sets (as it was at the time of the initial conjecture), there were, in fact, two obstacles encountered due to computing resources:

The power of Reconstructability Analysis (as implemented in *Occam3* as of 2011, and as running on a particular set of servers) is still insufficient to search the extremely large model space. This was not a solely matter of time—the program actually crashed in the case of some of the largest searches attempted.

The memory available on the lab computer or on campus student accounts was less than that needed to store the networks, test-result files, and learning-pass files for each of the 393 seed runs of the five networks in the final statistical comparison (3930 results files and 3930 network files). However, after deleting files to recover as much disk space as possible, sets of 388 files were able to be saved, which places the selection just below the minimum number required by Cohen's guidelines for statistical power. Moreover, those guidelines were developed for hand analysis of social data, and only serve as approximate guidelines for computational analysis of computational data. Since social data varies far more widely than computational data, the issue of drive space and the remaining five trials per network type was not pursued further.

7.8 Possible Sources of Error

In automating neural-net experiments for statistical significance, some control was lost over the exact behavior of NeuralWorks. For example, the Min/Max table's activation limits could only be set (or read from a file) when opening and setting up networks individually. The *automate.dat* file's list of commands for the *-xnautomate* feature did not include the Min/Max settings. (See Appendices K and L.) These settings protect the network being trained from paralysis by keeping activations away from the upper and lower ends of the range by a cushion (typically 20% of the activation range).

Not being able to have these cushions may have led to errors in two ways. First, network paralysis was not guarded against, and even if it did not occur, the details of the training of the random-seed replication experiments differed slightly from those of the preliminary runs. Secondly, in order to compare network performance “apples to apples,” the Min/Max limitation was imposed on the network outputs *a posteriori* at the thresholding stage. The arithmetic effect of subtracting 0.2 (or adding 0.2) *after the fact* versus that of limiting outputs to 0.2 below the maximum (or 0.2 above the minimum) *during training* must be somewhat different, but because of the numbers of experiments and the numbers of vectors used in the training and test sets, any errors introduced by thresholding after the fact will be distributed more-or-less equally among the different candidate network types and among the vectors (and thus output categories), the difference coming out to a wash (or a negligible offset) once the results are compared—a relative measure, rather than absolute. Hence, the final network-comparison results are deemed reliable.

Also potentially of concern is independence among the data sets as used for exercising different neural-net types. On one hand, we want the same data to be used for training (and same data for testing) each type of network so that a fair comparison can be made of the generalization under each seed. On the other hand, this implies that resulting performance data are not independent, and the results of statistical tests are suspect. As Salzberg observes, “even statisticians have difficulty agreeing on the correct framework for hypothesis testing in complex experimental designs” [33, p. 321], and that it may be preferable to use “random, distinct samples of the data to test each algorithm” [Ibid.]. In a multi-way comparison, the latter course of action would lead to each experiment being limited to a small fraction of the available data, which is problematic in its own right (though potentially alleviated through cross-validation). Furthermore, Prechelt maintains that: “[a]n algorithm evaluation is called acceptable if it uses a minimum of two real or realistic problems[,] and compares the results to those of at least one alternative algorithm” [136, p. 6]. The former would translate to an expansion of the current study to use the remaining teacher models in parallel studies. The latter has been done with the initial comparison (Table 8) with two RA generalizations and two human generalizations, but these trials were not repeated.

7.9 Shortcomings of the Dissertation

Comparison with other algorithms was limited to the preliminary experiments wherein four types of Reconstructability Analysis (RA) approaches were compared with the four types of prestructured neural nets (and the fully connected net, and human agents). These comparisons were against the use of RA as a classification (not just

modeling) tool. Since RA classification is not a matter of starting from a seed value, statistical comparison of RA classification with NN classification was not possible, but a different experimental design can be conceived where RA-search heuristics are somehow matched against NN-design heuristics, and a statistical comparison may be possible.

Another potential area for improving the multiple-comparison aspect of prestructuring research is to compare the generalization performances reported here with those of Self-Organizing Maps (SOM), Support-Vector Machines (SVM), various kernel methods (neural or otherwise), Ashenhurst-Curtis Logic Decomposition, and any other methods of Computational Intelligence and Statistical Learning.

Another major concern is that only one problem domain was utilized in the present study. A more thorough investigation of prestructuring for generalization improvement would apply the techniques to more than one problem domain or data source.

7.10 Strengths of the Dissertation

- Even though not statistically extensive, comparisons *were* made with another algorithm (RA classification) as well as two groups of human agents.
- The sample size for statistical (random-seeding) experiments was excessive rather than insufficient.
- The research undertaken was a blind venture, meaning domain characteristics were not known at the outset.

- Clave direction turned out to possess structure (as conjectured).
- RA was able to find non-intuitive clave-direction structure (as conjectured).
- The RA-prestructured networks improved on generalization (as conjectured).
- Such generalization improvement was statistically significant, with tight bounds, and mostly medium and large effect sizes.
- An extension¹⁰⁰ and verification of RA methodology for NN design is presented.
- A contribution is also made to the music-theory and music-education communities in the form of a musical theory to begin to help explain Afro-Brazilian principles of timing.
- A basis is provided for products in music education and music technology, as outlined in the researcher's *Journal of Music, Technology and Education* article adopted as Chapter V.
- And last but not least, intuitive tutorials on clave direction, Information Theory, Reconstructability Analysis, cross-validation and, of a sort, hypothesis testing are provided, with potential to be used as teaching materials.

7.11 Summary

In conclusion, it was shown that prestructuring using Reconstructability Analysis (as well as other means) can lead to improvement of artificial neural networks'

¹⁰⁰ To the case of noisy, real-world (even cultural) blind-venture cases from the already-structured, manufactured problem domains.

generalization performance in a classification task, and that such improvement can be statistically significant over that of the standard approach (the fully connected MLP), and moreover, that there are opportunities for future studies for solidifying these results to obtain greater confidence (or falsification) in the applicability of Occam's Razor through the Lendaris–Stanley conjecture to Neural Networks.

REFERENCES

- [1] Cook, G., *Teaching Percussion*, New York, NY: Schirmer Books, A Division of Macmillan, Inc., 1988.
- [2] Cruz, T., Moore, K., Gerald, M., and Fiol, O., *The Tomás Cruz Conga Method, Volume II—Intermediate—Essential Cuban Conga Rhythms*, Pacific, MO: Mel Bay Publications Inc, 2004.
- [3] Herder, R., *Spanish Montunos—Ensemble Studies in Rhythmic Precision and Pitch Security*, New York, NY: Continuo Music Press, Inc., 1972.
- [4] Lehmann, B., *The Syntax of ‘Clave’ – Perception and Analysis of Meter in Cuban and African Music*, Tufts University (Master’s thesis), Arlington, Massachusetts, 2002.
- [5] Machado, H. “Foca,” Muñoz, W., and Sadi, J., *El toque de Candombe—Ritmo tradicional Afro-Uruguayo/Traditional Afro-Uruguayan Rhythm*, Pacific, MO: Mel Bay Publications, Inc., 2002, pp. 10, 13, 22.
- [6] Mauleón, R., *Salsa Guidebook for Piano & Ensemble*, Petaluma, CA: Sher Music, 1993, pp. 48–57.
- [7] Mauleón-Santana, R., *101 Montunos*, Petaluma, CA: Sher Music, 2005.
- [8] Rodriguez, A., *Traditional Afro-Cuban Concepts in Contemporary Music* (Ethnic Percussion, ser. 3), Pacific, MO: Mel Bay Publications, Inc., 2003, pp. 41–45.
- [9] Sublette, N., *Cuba and Its Music: From the First Drums to the Mambo*, Chicago IL: Review Press, 2004.
- [10] Tobias, D. A., “Latin-American Percussion,” (The Instrumentalist, Nov. 1965) in *Percussion Anthology – A Compendium of Percussion Articles from The Instrumentalist Fourth Ed., including articles from 1946 through 1987*, Northfield, IL: The Instrumentalist Co., 1988.
- [11] Toussaint, G. T., “A mathematical analysis of African, Brazilian, and Cuban Clave Rhythms,” *Proceedings of BRIDGES: Mathematical Connections in Art, Music and Science*, 2002, pp. 157–168.
- [12] Novotney, E. D., *The 3-2 Relationship as the Foundation of Timelines in West African Musics*, University of Illinois at Urbana-Champaign (Ph.D. dissertation), Urbana-Champaign, Illinois, 1998.
- [13] Dawkins, R., *The Greatest Show on Earth: The Evidence for Evolution*, New York, NY: Free Press, 2009.
- [14] Allman, W. F., *Apprentices of Wonder: Inside the Neural Network Revolution*, New York, NY: Bantam Books, 1989.

- [15] Hoffman, R. E., Grasemann, U., Gueorguieva, R., Quinlan, D., Lane, D., and Miikkulainen, R., "Using computational patients to evaluate illness mechanisms in schizophrenia." *Biological Psychiatry*, Vol. 69, No. 10, 2011, pp. 997-1005.
- [16] Carroll, R. T., *The Skeptic's Dictionary: A Collection of Strange Beliefs, Amusing Delusions, and Dangerous Deceptions*, Hoboken, New Jersey: John Wiley & Sons, 2003.
- [17] Carey, T. V. (2011, Apr/May). "Parsimony (In as few words as possible)," *Philosophy Now*, Available:
www.philosophynow.org/issue81/Parsimony_In_as_few_words_as_possible
- [18] Kapur, J. N. and Kesavan, H. K., *Entropy Optimization Principles with Applications*, Academic Press, San Diego, 1992.
- [19] Kahneman, D., and Tversky, A., "Judgments of and by representativeness", in *Judgment under uncertainty: Heuristics and biases*, (Kahneman, D., Slovic, P., and Tversky, A. (eds), Cambridge: Cambridge University Press, 1982.
- [20] Fost, J., *If Not God, Then What?: Neuroscience, Aesthetics, and the Origins of the Transcendent*, Clearhead Studios, 2007.
- [21] Crick, F., *The Astonishing Hypothesis: The Scientific Search for the Soul*, New York, NY: Scribner, 1994, p. 57.
- [22] Lendaris, G. G., and Stanley, G. L., "Self-Organization: Meaning and Means," *Proc. 2nd Congress on Inf. Syst. Sci.*, Baltimore, MD: Spartan Books, 1965.
- [23] Börs, S., "How to Partition Examples between Cross-Validation Set and Training Set?," Laboratory for Information Representation, RIKEN, Japan, 1995.
- [24] Fiesler, E., and Beale, R. (eds), *Handbook of Neural Computation*, New York/Oxford: Oxford University Press, and Bristol/Philadelphia: Institute of Physics Publishing, 1997.
- [25] Bellman, R. E., *Adaptive Control Processes: A Guided Tour*, Princeton, NJ: Princeton University Press, 1961, p. 22.
- [26] S. Haykin, *Neural Networks – A Comprehensive Foundation (Second Edition/Low-Price Edition)*, Delhi, India: Pearson Education, Inc. (Singapore), 2004, p. 55.
- [27] Box, G. E. P., Hunter, W. G., and Hunter, J. S., *Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building*, New York, NY: John Wiley & Sons, 1978, p. 109.

- [28] Wolpert, D. H., and Macready, W. G., *No Free Lunch Theorems for Search*, The Santa Fe Institute, 1995, p. 24.
Available: www.santafe.edu/media/workingpapers/95-02-010.pdf
- [29] Leen, T., Slides 1–13, Artificial Neural Networks – 4, in course: (Advanced Topics in) Machine Learning, PSU CS 410/510/OGI CSEE 559/659, March 31, 2005.
- [30] Willett, K., and Zwick, M., “A Software Architecture for Reconstructability Analysis.” *Kybernetes*, Vol. 33, No. 5/6, 2004, pp. 997-1008.
- [31] Blum, A., Kalai A. Y., and Langford, J., “Beating the Holdout: Bounds for k-fold and progressive cross-validation,” in *Computational Learning Theory*, 1999, pp. 203–208.
- [32] Cavanaugh, J. E., “A Large-Sample Model Selection Criterion Based on Kullback's Symmetric Divergence,” *Statistics & Probability Letters*, Vol. 44, 1999, pp. 333-344
- [33] Kuha, J., “AIC and BIC: Comparisons of Assumptions and Performance,” *Sociological Methods & Research*, No. 33, 2004, pp. 188–229.
- [34] Salzberg, S. L., “On Comparing Classifiers: Pitfalls to Avoid and a Recommended Approach,” *Data Mining and Knowledge Discovery*, Vol. 1, 1997, pp. 317–328.
- [35] Scheffer, T., “Predicting the Generalization Performance of Cross Validatory Model Selection Criteria,” *Proc. 17th Intl. Conf. Machine Learning*, Stanford 2000.
- [36] Shibata, R., “Bootstrap estimate of Kullback-Leibler information for model selection,” *Statistica Sinica* No. 7, 1997, pp. 375-394.
- [37] NeuralWare NeuralWorks, *General Supplement*, p. 20.
- [38] Claeskens, G., and Hjort, N. L., “The focussed information criterion,” *Journal of the American Statistical Association*, No. 98, 2003, pp. 900–945.
- [39] Hastie, T., Tibshirani, R., and Friedman, J., *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (Springer Series in Statistics), New York, NY: Springer-Verlag, 2001, pp. 222, 241–249.
- [40] Rodriguez, C., “The ABC of Model Selection: AIC, BIC, and the new CIC,” *Bayesian Inference and Maximum Entropy Methods in Science and Engineering, AIP Conf. Proc.*, Vol. 803, 2005, pp. 80–87.
- [41] Stone, M., “An asymptotic equivalence of choice of model by cross validation and Akaike's information criterion,” *J. Roy. Statist. Soc. Ser. B*, No. 39, 1977, pp. 44–47.

- [42] Zhang, P., “Model Selection via Multifold Cross Validation,” *Annals of Stat.*, Vol. 21, No. 1, 1993, pp. 299–313.
- [43] Kohavi, R., “A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection,” *IJCAI’95*, pp. 1137–1143.
- [44] Busemeyer, J. R. and Wang, Y.-M., “Model Comparisons and Model Selections Based on Generalization Criterion Methodology,” *Journal of Mathematical Psychology*, Vol. 44, , 2000, pp. 171–189.
- [45] Hurvich, C., “The Corrected AIC (AIC_C),” [academic-course handout]. Available: <http://pages.stern.nyu.edu/~churvich/Forecasting/Handouts/AICC.pdf> Last accessed 5:44 PM, May 13, 2011
- [46] Krippendorff, K., *Information Theory: Structural Models for Qualitative Data*, (Quantitative Applications in the Social Sciences #62), Beverly Hills, CA: Sage, 1986.
- [47] Zwick, M., “4-Variable Structures,” Available: <http://www.sysc.pdx.edu/download/papers/lat4spec.pdf> Accessed 2007; last accessed 5:52 PM, May 13, 2011.
- [48] Shervais, S. and Zwick, M., “Using Reconstructability Analysis to Select Input Variables for Artificial Neural Networks,” *Proc. International Joint Conference on Neural Nets, Vol. 4*, 2003, pp. 3017–3021.
- [49] Vianna, H., *The Mystery of Samba: Popular Music & National Identity in Brazil*, The University of North Carolina Press, 1999.
- [50] Seppänen, J., “Tatum grid analysis of musical signals,” *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 2001, pp. 131–134.
- [51] Gómez, F., Khoury, I., Kienzle, J., McLeish, E., Toussaint, G., Melvin, A., Pérez-Fernández, R., and Rappaport, D., “Mathematical models for binarization and ternarization of musical rhythms,” *Proc. BRIDGES: Mathematical Connections in Art, Music and Science*, 2007, pp. 99–108.
- [52] London, J., *Hearing In Time: Psychological Aspects of Musical Meter*, New York, NY: Oxford University Press, 2004.
- [53] Danielsen, A., *Presence and Pleasure: The Funk Grooves of James Brown and Parliament*, Middletown, CT: Wesleyan University Press, 2006.
- [54] Toussaint, G., “Mathematical Features for Recognizing Preference in Sub-Saharan African Traditional Rhythm Timelines,” *Lecture Notes in Computer Science 3686*, Berlin/Heidelberg: Springer, 2005, pp. 18–27.

- [55] Kauffman, R., "African Rhythm: A Reassessment," *Ethnomusicology* Vol. 24, No. 3, 1980, pp. 393–415.
- [56] Kerman, J., *LISTEN: Brief Edition*, New York, NY: Worth Publishers, Inc., 1987, p. 20.
- [57] Hule, G., *Meter in Music, 1600–1800: Performance, Perception, and Notation*, Bloomington, IN: Indiana University Press, 1999.
- [58] Gerou, T., and Lusk, L., *Essential Dictionary of MUSIC NOTATION: The most practical and concise source for music notation is perfect for all musicians—amateur to professional*, Van Nuys, CA: Alfred Publishing Co., Inc., 1996.
- [59] Nikolov, K., private communication, May 15, 2011.
- [60] Johnston, D., *Weird Metronome*, 2007.
Available: www.pinkandaint.com/weirdmet.shtml (See the custom-entry option.)
- [61] Lamson, M., private communication, December 7, 2007.
- [62] Cohen, P. R., *Empirical Methods for Artificial Intelligence*, Bradford Books, 1995.
- [63] Lendaris, G. G., and Todd, D., "Use of Structured Problem Domain to Explore Development of Modularized Neural Networks," *Proceedings of International Joint Conference on Neural Networks '92 (IJCNN-92)*, Vol. III, 1992, pp. 869–874.
- [64] Lendaris, G. G., Zwick, M., and Mathia, K., "On Matching ANN Structure to Problem Domain Structure," *Proceedings of World Congress on Neural Networks'93 (WCNN-93)*, 1993, pp. 488–493.
- [65] Lendaris, G. G., and Mathia, K., "Using A Priori Knowledge to Prestructure ANNs," *Australian Journal of Intelligent Information Systems*, Vol. 1, No. 1, 1994, pp. 25–33.
- [66] Lendaris, G. G., and Mathia, K., "On Prestructuring ANNs Using A Priori Knowledge," *Proceedings of World Congress on Neural Networks '94 (WCNN-94)*, 1994, pp. 319–324.
- [67] Lendaris, G. G., Rest, A., and Misley, T., "Improving ANN Generalization Using A Priori Knowledge to Pre-Structure ANNs," *Proceedings of International Conference on Neural Networks '97 (ICNN '97)*, 1997, pp. 248–253.
- [68] Chambless, B., Lendaris, G., and Zwick, M., "An Information Theoretic Methodology for Prestructuring Neural Networks," *International Joint Conference on Neural Networks (IJCNN)*, 2001.
Available: www.sysc.pdx.edu/download/papers/nnra.pdf

- [69] Ripley, B. D., *Pattern Recognition and Neural Networks*, Cambridge, UK: Cambridge University Press, 1996.
- [70] Lammers, D., *IEEE Spectrum* (ISSN 0018-9235), New York, NY: The Institute of Electrical and Electronics Engineers, Inc., Vol. 47, No. 11 (NA), November 2010, p. 15.
Available: spectrum.ieee.org/semiconductors/processors/the-era-of-error-tolerant-computing
- [71] Dawkins, R., *Climbing Mount Improbable*, New York, NY: W W Norton & Company, 1997.
- [72] Mitchell, M., (class notes), PSU CS 410/510 TOP: Machine Learning, 2005.
- [73] Clark, A., *Being There: putting brain, body, and world together again*, Cambridge, MA: MIT Press, 1997, pp. 59–60.
- [74] Hand, D. J., “Pattern Detection and Discovery,” *Pattern Detection and Discovery, ESF Exploratory Workshop/Lecture Notes in Artificial Intelligence*, (Hand, D. J., Adams, N. M., and Bolton, R. J.; Eds.), Berlin: Springer-Verlag, 2002, pp. 1–9.
- [75] Clark, A., *Mindware: An Introduction to the Philosophy of Cognitive Science*, Oxford, UK: Oxford University Press, 2000, p. 19.
- [76] Borgida, A., and Sowa, J. F. (eds.), *Principles of Semantic Networks: Explorations in the Representation of Knowledge*, San Mateo, CA: Morgan Kaufmann Publishers, p. 3, 1991.
- [77] Abraham, T. H., “(Physio)logical circuits: The intellectual origins of the McCulloch-Pitts neural networks,” *Journal of the History of the Behavioral Sciences*, Vol. 38, No. 1, 2002, pp. 3–25.
- [78] McCulloch, W. S., and Pitts, W., “A logical calculus of the ideas immanent in nervous activity,” *Bulletin of Mathematical Biology*, Vol. 5, No. 4, 1943, pp. 115–33.
- [79] Rosenblatt, F., “The perceptron: A probabilistic model for information storage and organization in the brain,” *Psychological Review*, Vol. 65, No. 6, 1958, pp. 386–408.
- [80] Widrow, B., and Hoff, M. E., “Adaptive switching circuits,” New York, NY: *IRE Western Electric Show and Convention Record*, Vol. 4, 1960, pp. 96–104.
- [81] Mitchell, M., “Overview of Machine Learning,” (class notes), CS 510 Machine Learning, Portland State University, Winter Term, 2005, pp. 21–25.

- [82] Orr, G. (J.) B., “Linear Neural Networks”
Available: www.willamette.edu/~gorr/classes/cs449/linear2.html
Accessed 2:55 AM, May 12, 2008.
- [83] Minsky, M. L., and Papert, S., *Perceptrons: An Introduction to Computational Geometry*, Boston: MIT Press, 1969.
- [84] Hornik, K., Stinchcombe, M., and White, H., “Multilayer feedforward networks are universal approximators,” *Neural Networks*, Vol. 2, No. 5, 1989, pp. 359–66.
- [85] Lo, J. T.-H., “Multilayer Perceptrons and Radial Basis Functions are Universal Robust Approximators,” *Neural Networks Proceedings (IEEE World Congress on Computational Intelligence)*, Vol. 2, 1998, pp. 1311–4.
- [86] Minsky, M., *The Emotion Machine: Commonsense Thinking, Artificial Intelligence, and the Future of the Human Mind*, New York, NY: Simon & Schuster Paperbacks, 2006, pp. 289–91.
- [87] Grandvalet, Y., and Bengio, Y., “Semi-supervised Learning by Entropy Minimization,” *Advances in Neural Information Processing Systems*, Saul, L. K., Weiss, Y., and Bottou, L. (eds.), Cambridge, MA: MIT Press, 2005.
- [88] Zhu, X., “Semi-supervised Learning Literature Survey,” Technical Report 1530, Computer Sciences, University of Wisconsin-Madison.
Available: www.cs.wisc.edu/~jerryzhu/pub/ssl_survey.pdf, 2005.
- [89] Mann, G. S., and McCallum, A., “Efficient Computation of Entropy Gradient for Semi-Supervised Conditional Random Fields,” *Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics (HLT-NAACL); Companion Volume, Short Papers*, 2007, pp. 109–12.
- [90] Linsker, R., “Self-organization in a perceptual network,” *Computer*, Vol. 21, No. 3, 1988, pp. 105–117.
- [91] Fahlman, S. E., and Lebiere, C., “The cascade-correlation learning architecture,” in *Advances in Neural Information Processing Systems 2*, D. S. Touretzky, (ed.), San Mateo, CA: Morgan Kaufmann, 1990, pp. 524–532.
- [92] Albus, J. S., “A New Approach to Manipulator Control: The Cerebellar Model Articulation Controller (CMAC),” *J. Dyn. Sys., Meas., Control*, Vol. 97, No. 3, 1975, pp. 220–227.

- [93] Lang, R. I. W., *A Future for Dynamic Neural Networks*, Internal Report, University of Reading Cybernetic Intelligence Research Group (CIRG), 2000, p. 11.
- [94] Marchesi, M, Orlandi, G., Piazza, F., and Uncini, A., "An algorithm for Dynamically Adapting Neural Network Topologies," *Proceedings of the IEEE International Conference on Signals and Systems, ICCS '91*, 1991, pp. 56–60.
- [95] Parzen, E., "On estimation of a probability density function and mode," *Annals of Mathematical Statistics*, Vol. 33, 1962, pp. 1065-1076.
- [96] Wasserman, P. D., *Advanced Methods in Neural Computing*, New York, NY: Van Nostrand Reinhold, 1993, pp. 35–55.
- [97] Hinton, G. E., and Anderson, J. A. (eds.), *Parallel Models of Associative Memory*, Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers, 1981.
- [98] Hawkins, J., and Blakeslee, S., *On Intelligence*, Macmillan, 2005.
- [99] Hecht-Nielsen, R., "Counterpropagation networks," *Applied Optics*, Vol. 26, No. 23, 1987, pp. 4979–4983.
- [100] Grossberg, S., "Competitive Learning: From Interactive Activation to Adaptive Resonance," *Cognitive Science*, Vol. 11, 1987, pp. 23–63.
- [101] Carpenter, G.A. & Grossberg, S., "Adaptive Resonance Theory," In M. A. Arbib (Ed.), *The Handbook of Brain Theory and Neural Networks, Second Edition*, Cambridge, MA: MIT Press, 2003, pp. 87-90.
- [102] Zwick, M., "Wholes and Parts in General Systems Methodology," *The Character Concept in Evolutionary Biology*, G. Wagner, (ed.), Academic Press, 2001, pp. 237–256.
- [103] Anku, W., "Towards a Cross-Cultural Theory of Rhythm in African Drumming," *Intercultural Music*, Vol. 1, 1995, pp. 167–202.
- [104] Klir, G., *The Architecture of Systems Problem Solving*, New York, NY: Plenum Press, 1985.
- [105] Zwick, M., "An Overview of Reconstructability Analysis," *Kybernetes*, Vol. 33, No. 5/6, 2004, pp. 877–905.
- [106] Johnson, M. S., "State-Based Modeling Theory, Implementation and Applications," presentation, April 5, 2007, p. 5.
- [107] Zwick, M., (class notes), Discrete Multivariate Modeling, PSU SySc 551/651 January 15, 2008.

- [108] Ashby, W. R., "Constraint Analysis of Many-Dimensional Relations," *General Systems Yearbook*, No. 9, 1964, pp. 99–105.
- [109] Ashby, W. R., "Constraint Analysis of Many-Dimensional Relations," *Progress in Biocybernetics*, No. 2, 1965, pp. 10–18.
- [110] Klir, G. J., "Reconstructability Analysis: An Offspring of Ashby's Constraint Theory," *Systems Research*, Vol. 3, No. 4, 1986, pp. 267–271.
- [111] Klir, G. J., "The Role of Reconstructability Analysis in Social Science Research," *Mathematical Social Sciences*, No. 12, 1986, pp. 205–225.
- [112] Albright, S. C., Winston, W. L., and Zappe, C., *Managerial Statistics*, Monterey, CA: Brooks/Cole Publishing Company, 2001.
- [113] Carver, R. P., "The Case Against Statistical Significance Testing," *Harvard Educational Review*, Vol. 48, No. 3, 1978, pp. 378–99.
- [114] *Occam3* Manual.
Available: www.sysc.pdx.edu/download/papers/woccaman.pdf
- [115] Levitin, J., *This Is Your Brain on Music: The Science of a Human Obsession*, New York, NY: Dutton/Penguin Group (USA) Inc., 2006, p. 117.
- [116] Ekwueme, L. E. N., "Concepts of African Musical Theory," *Journal of Black Studies*, Vol. 5, No. 1, 1974, pp. 35–64.
- [117] Lendaris, G. G., private communication (e-mail), March 23, 2007.
- [118] Hinton, G. E., "Learning Distributed Representations of Concepts," *Proc. Eighth Annual Conf. of the Cognitive Science Society*, 1986, pp. 1–12.
- [119] Lendaris, G.G., "Experiment on Implementing the Concept-Type Lattice in a Neural Network," *Proc. AAAI Fourth Annual Workshop on Conceptual Graphs*, 1989.
- [120] Roberts, F. E., *Discrete Mathematical Models*, Prentice-Hall, Englewood Cliffs, NJ, 1976.
- [121] Warfield, J. N., *Societal Systems, Planning, Policy & Complexity*, Wiley-Interscience, New York, NY, 1976.
- [122] Lendaris, G. G., and Stanley, G. L., "Structure and Constraint in Discrete Adaptive Networks," *Proc. National Electronics Conference*, Vol. XXIIb, 1965, pp. 500–505.

- [123] Kamimura, R., Takagi, T., and Nakanishi, S., "Improving Generalization Performance by Information Minimization," *IEEE World Congress on Computational Intelligence*, Vol. 1, 1994, pp. 143–148.
- [124] Ragg, T., Gutjahr, S., and Sa, H. M., "Automatic determination of optimal network topologies based on information theory and evolution," *Proceedings of the 1997 23rd EUROMICRO Conference*, 1997, pp. 549–55.
- [125] Abraham, A., and Nath, B., "Hybrid heuristics for optimal design of artificial neural networks," *Third International Conference on Recent Advances in Soft Computing (RASC2000)*, 2000.
- [126] Ragg, T., Braun, H., and Landsberg, H., "A Comparative Study of Neural Network Optimization Techniques," *13th International Conference on Machine Learning: Workshop Proceedings on Evolutionary Computing and Machine Learning*, 1996, pp. 111–118.
- [127] Bianchini, M., and Gori, M., "Optimal learning in artificial neural networks: a theoretical view," *Proc. of Summer School on Rec. NN*, Italy, 1997.
- [128] Drucker, H., and Le Cun, Y., "Improving generalization performance using double backpropagation," *IEEE Transactions on Neural Networks*, Vol. 3, No. 6, 1992, pp. 991–997.
- [129] Hashem, S., "Optimal Linear Combinations of Neural Networks," *Proceedings of the IEEE Conference on Neural Networks*, Vol. 3, 1994, pp. 1507–12.
- [130] Lu, B.-L., and Ito, M., "Task Decomposition and Module Combination Based on Class Relations: A Modular Neural Network for Pattern Classification," *IEEE Transactions on Neural Networks*, Vol. 10, No. 5, 1999, pp. 1244–1256.
- [131] Freedman, K. B., and Bernstein, J., "Current Concepts Review - Sample Size and Statistical Power in Clinical Orthopaedic Research," *J Bone Joint Surg Am.*, No. 81, 1991, pp. 1454–60.
- [132] Siegfried, T., "Odds Are. It's Wrong: Science fails to face the shortcomings of statistics," *Science News – The Magazine of the Society for Science & the Public*, Vol. 177, No. 7, 2010, p. 26.
- [133] Ioannidis, J. P. A., "Why Most Published Research Findings Are False," *PLoS Medicine*, Vol. 22, No. 8, 2005, p. e124.
- [134] Miller, J., "Statistical significance testing—a panacea for software technology experiments?" *The Journal of Systems and Software*, 73, 2004, pp. 183–192.
- [135] Zucchini, W., "An introduction to model selection," *Journal of Mathematical Psychology*, No. 44, 2000, pp. 41–61.

- [136] Prechelt, L., "A quantitative study of experimental evaluations of neural network algorithms: Current research practice," *Neural Networks*, No. 9, 1996.
- [137] Flexer, A., "Statistical Evaluation of Neural Network Experiments: Minimum Requirements and Current Practice," In R. Trappl, ed., *Cybernetics and Systems '96: Proc. 13th European Meeting on Cybernetics and Systems Res.*, 1996, pp. 1005–1008.
- [138] Holte, R., "Very simple classification rules perform well on most commonly used datasets," *Machine Learning*, Vol. 11, No. 1, 1993, pp. 63–90.
- [139] Tversky, A., and Kahneman, D., "Belief in the law of small numbers," in D. Kahneman, P. Slovic, and A. Tversky (eds.), *Judgment Under Uncertainty: Heuristics and Biases*, Cambridge University Press, 1971, pp. 3–20.
- [140] Fraisse, P., *The Psychology of Time*, trans. By J. Leith, New York, NY: Harper & Row Publishers, 1963, p. 71, p. 72, and pp. 82–83.
- [141] Dubos, R., *Mirage of Health: Utopias, Progress, and Biological Change*, New Brunswick, NJ: Rutgers University Press, 1996, p. 168.
- [142] Bausell, R. B., *Snake Oil Science: The Truth About Complementary and Alternative Medicine*, Oxford, 2007, Chapters 1 & 2.
- [143] Freedman, K. B., Back, S., and Bernstein, J., "Sample size and statistical power of randomised, controlled trials in orthopaedics," *J Bone & Joint Surg Br*. Vol. 83-B, No. 3, 2001, p. 397.
- [144] Berger, V. W., PhD, and Hsieh, G., PhD, "Rethinking Statistics: Basing Efficacy Alpha Levels on Safety Data in Randomized Trials," *Israeli Journal of Emergency Medicine*, Vol. 5, No. 3, 2005, pp. 55–60.
- [145] Mlodinow, L., *The Drunkard's Walk: How Randomness Rules Our Lives*, New York, NY: Vintage Books, 2008.
- [146] Ziliak, S. T., and McCloskey, D. N., *The Cult of Statistical Significance: How the Standard Error Costs Us Jobs, Justice, and Lives*, Ann Arbor, MI: The University of Michigan Press, 2008.
- [147] Forster, M. R., "Key Concepts in Model Selection: Performance and Generalizability," *Journal of Mathematical Psychology*, 44, 2000, pp. 205–231.
- [148] Forster, M., and Sober, E., "How to Tell When Simpler, More Unified, or Less Ad Hoc Theories will Provide More Accurate Predictions," *Br J Philos Sci*, Vol. 45, No. 1, 1994, pp. 1–35.
- [149] Shao, J., "Model Selection by Cross-Validation" *Journal of the American Statistical Association*, Vol. 88, No. 422, pp. 486–494, 1993.

- [150] Rao, R., and Wu, Y., "A strongly consistent procedure for model selection in a regression problem," *Biometrika*, Vol. 76, No. 2, 1989, pp. 369–374.
- [151] Gould, S. J., *Full House: The Spread of Excellence from Plato to Darwin (Uncorrected Proof/ Special Advance Reader's Edition)*, New York, NY: Harmony Books, 1996.
- [152] Gelman, A., and Stern, H., "The Difference Between 'Significant' and 'Not Significant' is not Itself Statistically Significant," *The American Statistician*, Vol. 60, No. 4, 2006, pp. 328–331.
- [153] Baroni, M., "The Concept of Musical Grammar," *Music Analysis*, Vol. 2, No. 2, 1983, pp. 175–208.
- [154] Bamberger, J., *The Mind behind the Musical Ear: How Children Develop Musical Intelligence*, Cambridge, MA: Harvard University Press, 1991, p. 20.
- [155] Koetting, J., "Analysis and Notation of West African Drum Ensemble Music," *Selected Reports in Ethnomusicology*, Vol. 1, No. 3, 1970, pp. 115–46.
- [156] Cohen, J., "A Power Primer," *Psychological Bulletin*, Vol. 112, No. 1, 1992, pp. 155–159.
- [157] MacKay, D. J. C., *Bayesian methods for adaptive models*. Dissertation (Ph.D.), California Institute of Technology, 1992.
Available: resolver.caltech.edu/CaltechETD:etd-01042007-131447
- [158] Spiro, M., and Ryan, J., *The Conga Drummer's Guidebook*, Petaluma, California, USA:, Sher Music Co., 2006.
- [159] Wright, M., Schloss, W. A., and Tzanetakis, G., "Analyzing Afro-Cuban Rhythm Using Rotation-Aware Clave Template Matching with Dynamic Programming," *Proc. 9th International Conference on Music Information Retrieval*, 2008, pp. 647–652.
- [160] Washburne, C., "The clave of Jazz: A Caribbean Contribution to the Rhythmic Foundation of an African-American Music," *Black Music Research Journal*, Vol. 17, No. 1, 1997, pp. 59–80.
- [161] Gomes, S., *New Ways of Brazilian Drumming*, Tübingen, Germany: advance music, 2007.
- [162] Rosauero, N., *The ABCs of Brazilian Percussion*, Brazil: Propercussã, 2004, p. 7.
- [163] Flanagan, P., "Quantifying Metrical Ambiguity," *Proc. 9th International Conference on Music Information Retrieval*, 2008, pp. 635–640.

- [164] Collins, J., *African Musical Symbolism in Contemporary Perspective (Roots, Rhythms and Relativity)*, Berlin, Germany: Pro BUSINESS GmbH., 2005, p. 398.
- [165] Díaz-Báñez, J. M., Farigu, G., Gómez, F., Rappaport, D., and Toussaint, G. T., “El Compás Flamenco: A Phylogenetic Analysis,” *Proceedings of BRIDGES: Mathematical Connections in Art, Music and Science*, 2004, pp. 61–70.
- [166] Berliner, P., *Thinking in Jazz: The infinite art of improvisation*, Chicago, IL: The University of Chicago Press, 1994, p. 492.
- [167] Gonçalves, G. & Mestre Costa, O., *Aprendendo a Tocar O BATUQUE CARIOCA: As Baterias das Escolas de Samba do Rio De Janeiro*, Rio De Janeiro, Brazil: Groove, 2000.
- [168] Henderson, W. J., *What Is Good Music? Suggestions to Persons Desiring to Cultivate a Taste in Musical Art*, New York, NY, USA: Charles Scribner’s Sons, 1925, pp. 33–34.
- [169] Sloboda, J. A., *The Musical Mind: The cognitive psychology of music*, Oxford, England, UK: Clarendon Press, 1985.
- [170] Mello, M. (composer), as interpreted by Daniela Mercury (1997), ‘Feijão Com Arroz’, [CD], Sony Latin, CDZ-82118 2-479390.
- [171] McClave, J. T., and Sincich, T., *Statistics*, Upper Saddle River, NJ: Prentice Hall, 2003, p. 94.
- [172] Raudys, Š., *Statistical and Neural Classifiers: An Integrated Approach to Design (Advances in Pattern Recognition)*, London: Springer-Verlag, 2001, pp. 191–194.

LIST OF APPENDICES

Appendix A: A Cross-Cultural Grammar for Temporal Harmony in Afro-Latin Musics:

Clave, *Partido-Alto*, and Other Timelines

Appendix B: Information Theory

Appendix C: Literature Review of Music Information Retrieval and Digital (Audio)

Signal Processing

Appendix D: Statistics, Social Responsibility, and the Enhanced Scientific Method

Appendix E: The Data-Acquisition Process

Appendix F: Brazilian Glossary, Clave Terminology, Musicological Definitions, and

Samba Concepts

Appendix G: Northern/Non-Northern vs. Western/Non-Western

Appendix H: Occurrences of Clave-Type Patterns in Music around the World, and

Recommended Listening

Appendix I: Author's Musical Background and Qualifications

Appendix J: A Survey of Notational Conventions and Preferences for Meter and Time

Signature for Musics of the African Diaspora

Appendix K: The *automate.dat* Script for Neural-Net Experiments in NeuralWare's

NeuralWorks

Appendix L: The *AutomateNeuralWorks.vbs* Visual Basic Script for Exercising Network

Configurations with Multiple Seeds

Appendix M: C Code for the top-generalizing BIC-based network

Appendix N: Sample Data in the Firm-Teacher Context

Appendix O: Random Seeds

APPENDIX A: A Cross-Cultural Grammar for Temporal Harmony in Afro-Latin Musics: Clave, *Partido-Alto*, and Other Timelines

The relationship between the plane of expression and the plane of content ... is made possible only through the existence of socially and culturally accepted conventions—in semiotic terminology, a code. (Baroni, p. 181)

This appendix started out in 2001 as a clave tutorial in which the idea of clave direction was represented as a cultural code, and eventually grew to a proposed theory of clave (a grammar for temporal harmony). It has been submitted to a leading musicology journal, and the initial reviews are very encouraging. The version presented here is slightly improved.

The relevance of the material herein to both computer technology and music theory is reflected in another of Baroni's observations:

[C]omposition with a computer ... presupposes an analysis made in advance Musicologically, this becomes a significant procedure when applied to pastiche composition; here, the use of a computer *reveals latent structure* which enables the music to be reproduced. It is therefore *vital to make an analysis for a particular style* and eventually to *elaborate a theory to account for it* (Baroni, p. 179).

The steps identified by Baroni and emphasized in bold precisely parallel the history, objective, and contribution of this appendix.

A.1 Introduction

This article presents a formal study of the musical concept called *clave direction*. The significance and regulative role of clave and similar cyclical timelines¹⁰¹ in Afro

¹⁰¹ In many other countries, and among scholars worldwide, various other names and descriptions are in use to refer to the same fundamental principle. Alternate names include clips (Collins 2004: 31) and madera (Machado, Muñoz et al. 2002: 10, 13, 22). Further descriptive phrases favored by scholars to describe the duple- and triple-time clave-type patterns include 'the principle of mobility and finality' (Anku 1995: 172–3), 'generative concept' (Spiro & Ryan 2006: 7), 'rhythmic key' (Sublette 2004: 168), 'standard pattern' (Novotney, 1998: 116–121), 'timeline' (Ekwueme 1974: 61), 'structural device' (Mauleón-Santana 2005: 5), and 'key pattern' (Novotney 1998: 160–165).

Latin music are attested to by many scholars. For example, it is identified as the rhythmic anchor of Cuban music (Sublette 2004, 95, 166–167, 342–343; Cruz, Moore et al. 2004, 75; Soebbing 1988, 524; Cook 1988, 321–324; Mauleón-Santana 2005, 1, 7–9; Herder 1972, 41). Furthermore, some issues of terminology notwithstanding, musicians and scholars also agree that clave is the central regulating principle not only of Cuban, but of all Afro-American music (Gomes 2007: 90–91; Rosauero 2004, 7; Novotney 1998, 168, 236–238; Machado, Muñoz et al. 2002, 10, 13, 22; Washburne 1997, 59–60, 66–67).

Clave is not just a pattern but a critical concept. This is attested to by many educators (Rodríguez 2003, 41–45; Mauleón-Santana 2005, 5–8; Spiro and Ryan 2006, 12–17). Specifically, “clave beat is the foundation of Latin-American rhythm and practically all of the other instruments are guided by this beat” (Tobias 1965, 270); “clave is the key to understanding how Afro-Cuban music is arranged and flows” and “a concept that is fundamental to Afro-Latin music” (Rodríguez 2003, 41).

The word “clave” can take on several meanings¹⁰². The first three are well-known: (1) clave the instrument; (2) clave in the harmonic sense, as in “the key of C[#]” (in Spanish); and (3) a small family of patterns: the clave-proper (what musicians typically have in mind for clave). This article focuses on the fourth meaning of clave, which is frequently mentioned but rarely investigated. We refer to this as *wide-sense clave*: clave as a concept and a rhythmic-regulative principle. In this wider sense, clave-the-concept is not just the existence of specific patterns, but the relationships that any pattern may have with a family of associated patterns. This is what Malabe and Wiener mean when they observe that “[a]ny rhythmic figure can serve as a clave” (Malabe and Wiener 1990, 9). Similarly, as Sublette contends, “clave is not a beat ... [but] a key: a way of coordinating independent parts of a polyrhythmic texture.” (2004, 170–171).

The relationship between cyclical timelines that act as clave patterns and the specific role they play in their respective traditions is *precisely* analyzed and explained below, in a manner that unifies these musical traditions under one regulative concept (distilled from traditional practice as taught by masters and exemplified by recording artists). Previously, the clave concept has been analyzed and explained by Novotney (1998), Lehmann (2002), Toussaint (2002), and Mauleón-Santana (2005) in various valuable ways. The analysis presented here is novel because it differs in its *consistency* (from case to case), *parsimony* (only two concepts suffice) and *precision* (those two concepts are precisely defined)¹⁰³. As long as one can express rhythmic accent patterns

¹⁰² See Mauleón’s *101 Montunos* for a similar distinction that informed this terminology (2005).

¹⁰³ This grammar of clave is consistent and precise, but not complete: Certain simplifying assumptions had to be made to get this first-order grammar up and running. Further research is needed

as strings of 16 onsets (through quantizing and thresholding), a conclusive identification can be made as to the clave direction (or lack thereof) for any pattern. (A few outstanding issues resulting from certain simplifying assumptions are listed in the conclusion and in the end notes.)

The analysis of clave direction presented here is meant to be useful to the ethnomusicology, music-education, music-theory and Latin-music communities—that is, for listeners, performers, educators and researchers alike—by contributing to a broader understanding of clave while respecting its form and tradition. The present approach arose from a commitment to performing the music with care for and appreciation of the nuances of the music of the original cultures. The grounds for this approach and the ways such an analysis can benefit music analysis, music-making and the music industry is expounded in the present author’s corresponding article (2011).

The regulative role of clave is aided greatly by a metric called *offbeatness* (Toussaint 2005, 23), which is reinterpreted here into the idea of *relative offbeatness*¹⁰⁴. Through this key concept, the regulative role of clave can be explained simply and consistently. Parenthetically, we should note that clave also plays other organizational roles: those related to phrasing and time-keeping. The former is discussed in detail by Herder (1972) in terms of when to start and stop improvisational phrases, and also in (Mauleón 1993).

Wide-sense clave provides resolutions to common difficulties even professional musicians have encountered in reconciling the African-based, Cuban-developed clave concept with its function in other Afro-Latin musics¹⁰⁵. (See Sections 6.4 and 6.5.)

to expand what is presented here to a wider, practical coverage of dynamics, tempi, expressive timing and other time signatures.

¹⁰⁴ Relative offbeatness is a novel concept put forth by the present author (regarding the necessity of considering the entire rhythmic entity before coming to a decision about its clave direction), and supported by evidence (see Section 4). To know whether a given pattern features offbeatness for purposes of clave-direction identification, one must ask “relative to what other pattern?” An excellent hypothesis as to why clave direction exists in the first place, consider that “continual use of off-beating without respite would cause a readjustment on the part of the listener, resulting in a loss of the total effect; thus off-beating [with respite] is a device whereby the listeners’ orientation to a basic rhythmic pulse is threatened but never quite destroyed” (Merriam 1959, 16).

¹⁰⁵ For example, the 3-2 clave from Belize has two of the notes of the Cuban 3-2 son clave on its 3-side, and adds a second downbeat to the Cuban clave’s 2-side, resulting in two notes during what is obviously (by sound, feel and function) the 3-side, and three notes during the 2-side. Likewise, the typical third-surdo ostinato played in most samba batucada has two strikes in the first half of the phrase and three in the second half, yet is always played over 3-2 samba (because it is in 3-2). For the most part, samba musicians either don’t think about it, or have difficulty reconciling the fact that there are two onsets on the 3-side, and three onsets on the 2-side.

The clarifications offered below for understanding and applying clave direction under various circumstances, for various instruments, and to a variety of Latin American musics should help musicians at virtually any level of experience further their understanding and appreciation of clave. Nonetheless, this analytical approach is only offered as a theoretical tool for the clave-conscious musician's toolbox, not as an alternative to developing an intuitive feel for clave through practicing and listening.

As in any focused investigation of a complex phenomenon, this examination of clave is undertaken with a few *necessary assumptions*: (1) that quantized attack-point rhythm—which disregards durations, releases, and expressive timing (swing, rubato, etc.)—is sufficient for a *basic*, initial study of clave direction; (2) that in the interest of brevity, it is adequate to focus on a small number of styles (but pointing out their close ties with other traditions, as is done throughout this paper); and (3) that the time-keeping and phrasing aspects of clave are both less challenging and outside the scope of this treatment (leaving the offbeatness-regulative aspect of clave as the sole focus of the present article). Even though the analysis below has so far only been applied to some Cuban, Brazilian, Haitian, Belizean, Ghanaian¹⁰⁶ and US (funk) music, we expect the approach to be relevant to all Afro-Latin musics.

There are two intentional substitutions in this article that differ from the norm in the musicology literature: the use of the term *offbeatness* (in place of the more familiar *syncopation*¹⁰⁷) and the use of the Time-Unit Box System (TUBS: Koetting 1970, 115-146), also known as “matrix notation” (Rodriguez 2003, 63) in conjunction with standard music notation. Strong support for the use of TUBS is given by Koetting and Knight in their discussion of hemiola, hocket, and “fastest pulse” (Koetting and Knight 1986, 60). Hence, although European music notation is the standard for all written

Other difficulties in reconciling the naming of patterns versus their function include *partido-alto* with its seven onsets, Cuban *cáscara* with its ten onsets, and the 3-2 *rumba* clave, considered 2-3 in practice.

¹⁰⁶ While this article focuses much more on samba due to the author's expertise in that style, Kauffman's transcription of the Ewe dance *Sohu*, (1980, 411) and Kwabena Nketia's transcription of the Akan songs *Kwasi Dente* (1970, 132) and *Akatapefoɔ* (145), and the Mamprusi song *Darbɔrga* (141) are in keeping with the idea of clave direction as described here.

¹⁰⁷ This usage arose as a result of research into the cultural dimensions of the term syncopation: Syncopation is norm-referenced (subjective), meaning accents fall in non-normal places (Kauffman 1980, 394; Kerman 1987, 20; Novotney 1998, 104 and 108), while offbeatness is tactus-referenced (objective). Readers not wishing to be nitpicky can use either expression.

music, TUBS¹⁰⁸ is included here because of the correspondence between the geometrical positions of notated onsets and the timing of note attacks (Figure 1).

Lastly, the analysis offered in this article is not intended to impose an etic analysis on these art forms, or to claim that clave can be reduced to numbers. On the contrary, the intent is to contribute to a broader understanding of clave while respecting its form and tradition. Although the way to internalize and *feel* clave is to listen, practice, and play in clave-aware settings, the analytically oriented reader may still benefit greatly from the detailed analysis offered here.

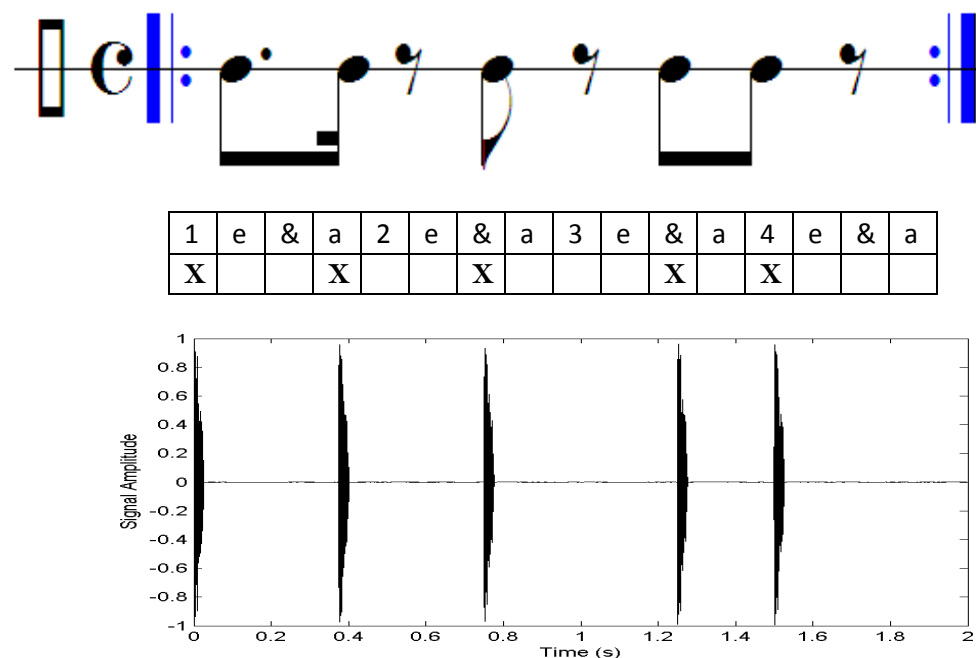


Figure 1: The 3-2 *son* clave in standard notation, TUBS, and sound waveform. Note the visual correspondence between the waveform and TUBS. Careful spacing is required in standard notation to get the same correspondence. (The first line of the TUBS representation shows one way to vocalize sixteenth-note subdivisions, pronounced “one, ee, and, ah.”)

¹⁰⁸ TUBS can be a helpful abstraction that both beginners and trained musicians can understand. For beginners, it is more easily understood than standard notation, but trained musicians (who can express themselves in standard notation when necessary) also find it easy to conceive of rhythms in TUBS. Rests, rolls, and sustained tones are mostly irrelevant for the initial study of clave. For clarification, see Locke’s discussion on dotting, tying, and rests (1987, 9–10).

A.2 The Basic Claves (Clave-Propre)

A.2.1 The (Afro-Cuban) *Son* Clave: A seemingly obvious starting point

Son clave is a form that appears in many music styles worldwide (Figure 1). If we took the standard course in the discussion of the clave, and break it up into the first half and second half of the phrase, examining each half in terms of syncopation and resolution, we would find three onsets in one half and two onsets in the other. (In fact, it was after these numbers that the clave directions were named). While making perfect sense in terms of the *son* clave, this approach does not engender a conception of other patterns which are traditionally understood (heard) to be in *clave agreement* with the *son*.

We will refer to this typical method of counting the onsets in the first and second halves of the phrase as the “standard approach,” and show how the novel approach of relative offbeatness supplants and surpasses it. Although the *son* clave lends itself easily to the standard approach, experience with many other rhythms in their traditional contexts shows that analysis based on relative offbeatness works (in terms of agreeing with the respective traditional practices) in more clave settings¹⁰⁹.

In the new method, we count onsets according to their rhythmic function, and instead of an absolute count (three onsets here, two there), we use a relative count: more onsets supporting this function, fewer supporting that function. (See the discussion of *surdo de tercera* in Section A.2.6.4 for the relevance of this). Instead of first half/second half, we look at the inner part and the outer part of the pattern (Figure 2). And instead of syncopation and resolution, we consider offbeatness and onbeatness. Unlike the standard approach, the inner/outer perspective combined with relative offbeatness works in virtually every case from Ghanaian *gahu* (Toussaint 2002) to American funk (see section 6.2), and from Haitian *konpa* (section 6.1) to Brazilian *axé*.

¹⁰⁹ The standard approach in fact demonstrates two of the problems solved with the new approach: The way this pattern starts right on the first downbeat (the one) may give the impression that the first half of the pattern is less syncopated than the second. This is one of the practical reasons offbeatness is a more appropriate concept for clave than syncopation. Examining the first half of the pattern in terms of inter-onset intervals (IOIs, London, 2004), one finds that there is syncopation in the relationship of the notes in the first half, but not in the second. So, is the first half more syncopated or less syncopated? It depends on what level of the rhythmic hierarchy we look at, and whether we consider IOIs or the onsets themselves. Aside from not being a problem in the clave analysis presented here, this apparent irony is part of the ingenuity of the African and Afro-Latin forms. Since the less-offbeat section does not start on a downbeat at all, the two sections maintain separate senses of syncopation at different levels of the hierarchy (which do not constitute different cases of offbeatness, however).

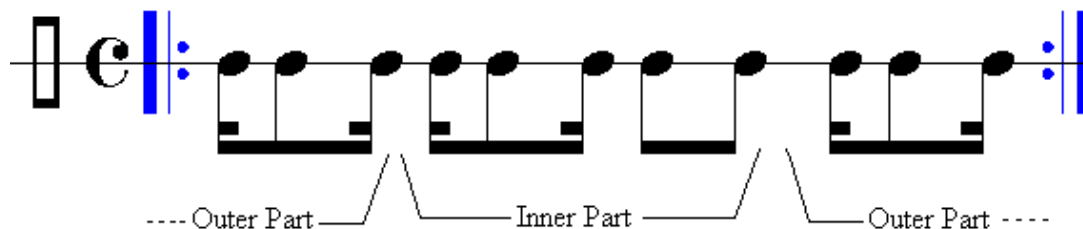


Figure 2: The inner and outer parts of the *okele* (phrase cycle, Ekwueme 1974, 46), depicted using the *cáscara*.

Let's define the *inner part* of the phrase as ranging from the downbeat of *two*, up to the downbeat of *four*. The *outer part*, then, ranges from the downbeat of *four*, around to the downbeat of *two*. (This overly simple definition is loosened in Section 5.4.) Let's use a naming scheme for the subdivisions of the quarter note (Figure 3) such that both downbeats and upbeats are considered *onbeats*.

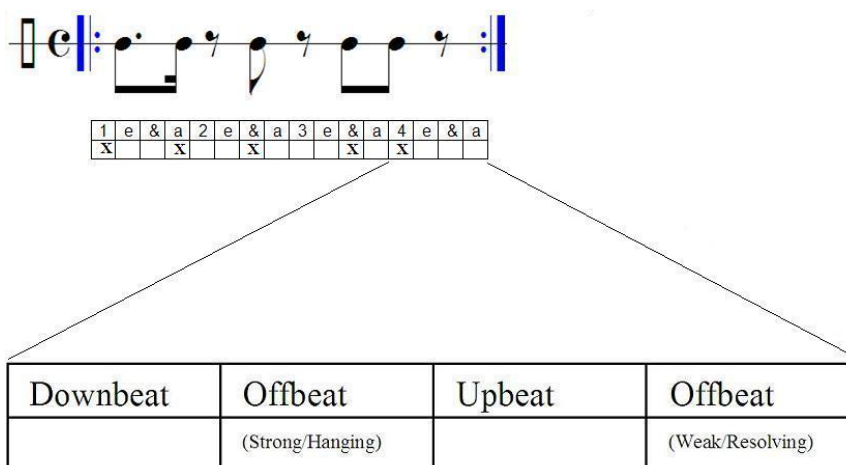


Figure 3: Suggested naming scheme for sixteenth-note subdivisions of the quarter note in Afro-Latin music.

The inner part of the 3-2 *son* clave consists solely of two onbeats while the outer part consists of two onbeats as well as one offbeat. Since this pattern is typically called “3-2,” we will refer to any other pattern exhibiting this offbeatness behavior (more *off* outside, more *on* inside) as 3-2. This label will apply even when two onsets show up in the first half, or three in the second half, as long as relative offbeatness follows this

arrangement. The reasons for this scheme will become clearer as we examine other patterns.

It is important to reiterate that although speaking of a neat division into first and second halves would have helped us understand the *son* clave just as well, the standard approach fails with many of the more complex patterns found in Afro-Latin musics.

What makes the *son* clave easy to model as a tension half and a resolution half are two characteristics: 1) The particularly neat distribution of inter-onset intervals (time between note attacks); 2) the complete lack of onsets from beat four to beat one. This rare situation gives rise to potentially misleading mental models for claves. The latter characteristic, as we will shortly see, is not the case for the so-called “*bossa*” clave, which is a more difficult pattern (Toussaint 2002, 4–7) that is generally considered vague in terms of explicating clave direction. On the contrary, the offbeatness-based analysis presented here sheds light on the *bossa* clave as an even more revealing pattern than *son* clave for indicating clave direction.

A.2.2 The So-Called Brazilian (*Bossa*) Clave

Compared with the *son* and the *rumba* clave (for the latter, see Rodriguez 2003, 48; Tomás, Moore et al. 2004, 75; Spiro and Ryan 2006, 13), this clave has very different characteristics, even though it may look and sound like the others. As a result of this difference, however, it provides an excellent demonstration of the wider meaning of clave direction. Offbeatness and onbeatness exhibit themselves (and then expire) earlier than the first-half/second-half notion would imply. The “3-side” really begins *before* the first half of the phrase, and ends sooner than the midpoint of the phrase. Likewise, the 2-side begins and ends sooner than the second half.

This pattern (sometimes called *bossa* clave) is widely used in various Brazilian styles. It indicates the *partido-alto*—the true key to samba—more closely than the *son* clave. Though controversy exists over whether there is clave in Brazilian music (Rosauero 2004, 7; Gomes 2007, 88; Machado, Muñoz et al. 2002, 10), samba is indeed played within the *partido-alto* form (analyzed below). Since musicians are expected to know, feel, and play within the underlying rhythmic structure of each Brazilian form, Afro-Brazilian music evidently does follow the clave concept¹¹⁰.

¹¹⁰ However, we must note that in lusophone Brazil, the word clave literally means ‘clef’, and the equivalent of the Spanish-language clave (key) is *chave* (pron. sha-vee). One can avoid the

To arrive at the bossa clave (Figure 4), delay the last onset of the *son* clave (Figure 1) by one subdivision. Note that the sense of finality found in the Cuban *son* clave vanishes in the *wrap-around* feel of the Brazilian pattern. Inter-onset intervals in each half of the bossa clave are the same, so the distribution of onbeat and offbeat onsets is more even than in the Cuban claves. Hence speaking of the first and second halves gets us nowhere in terms of tension and resolution typically used to analyze the Cuban claves.

If we consider the demarcation shown in Figure 2 instead, we find that the outer part only has one onset on a downbeat, and two onsets on offbeats (the *a* of *one*, and the *e* of *four*). The inner part has the same two onbeats as the *son* clave. Hence, the pattern has greater outer offbeatness (or, greater inner onbeatness) consistent with the label of 3-2. It is no longer a problem that the three-side of either clave has the main downbeat and the two-side does not¹¹¹. This is partly because downbeats and upbeats contribute roughly the same sense of onbeatness to the overall pattern. (There are, of course, context-dependent differences, but in this first approximation to clave, we can overlook those without harm.)

In finding the direction of the *bossa* clave, it is not sufficient to count three onsets in the first half and two in the second half because dozens of examples like Figure 4, bottom, can be generated with three onsets on the “two side” and two onsets on the “three side.” Besides, dividing the phrase in two and counting note onsets cannot be sufficient to identify implied direction because clave, as attested to in all sources, governs the timing and phrasing of *all parts* and *all instruments*. Clearly, not all parts can be limited to five notes per phrase. It is the local variations in offbeatness that determine (and are determined by) clave direction, not whether two or three note attacks are being played. Hence, resolving the overall feel produced by any pattern requires intuiting how the rhythms of interacting instruments will relate to one another, *not* counting notes. The standard approach would give the same clave direction (3-2) for the patterns in Figure 4. But the bottom figure is considered “crossed” (in the wrong clave direction) if used as an ostinato in any 3-2 instance of Afro-Latin music, Cuban, Brazilian or otherwise. The wide-sense clave approach correctly identifies clave direction in these and myriad other patterns.

contentious term clave (Cruz, Moore et al. 2004, 77–80) by favoring either *chave* or *partido-alto* when speaking of samba.

¹¹¹ This clave problem was posed to the author by M. Spiro at California Brazil Camp 2005.

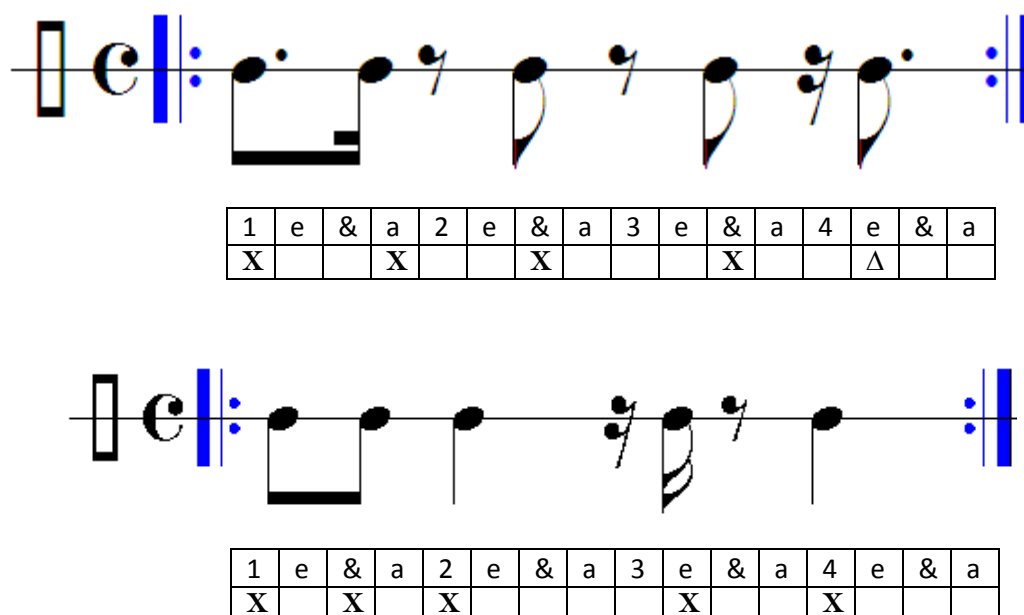


Figure 4, top: The 3-2 bossa clave. The onset shown with a triangle indicates how the last note of *son* can be delayed to obtain this form. Bottom: An arbitrary pattern in 2-3, illustrating that counting onsets does not always indicate clave sense.

When approaches and interpretations of clave direction differ, the final arbiter ought to be the accepted practice in the tradition in question¹¹². When superimposed with other rhythms that are more apparent in their clave direction (such as *partido-alto*), the *bossa* clave produces a tangible sense of *resistance* for experienced musicians when played over a 2-3 pattern, but fits comfortably over a 3-2 pattern. Hence, it is 3-2.

A.2.3 A Graphical Analogy for Clave Direction

Before moving on to other patterns, a graphical analogy is used below to present the idea in an alternate manner. Imagine that in a fictional culture, visual decorations, no matter how elaborate, generally follow a pattern of alternating round and pointy designs. Assume for the sake of argument that stars and circles have traditionally been used to provide the alternating pattern, and that the arrangement in Figure 5, line one, is called the ‘forward’ pattern. A pattern with circles first and then stars would be ‘reverse’.

¹¹² This may, of course, be difficult to establish. One may query keepers of the tradition, live within the culture, or learn from the masters. Along with the first and last of these options, the analysis presented in this article also stems from listening trials conducted with master drummers.

Now imagine that an artisan wishes to expand his options beyond stars and circles, and introduces a rhombus. However, wishing to remain true to tradition, s/he must consider whether a rhombus is more or less pointy than a star or a circle. The artisan concludes that a rhombus is less pointy than a star (Figure 5, line two), but more pointy than a circle (Figure 5, line three).

The artisan has managed to stay within tradition while introducing new elements. If other elements such as triangles or even arbitrary shapes are introduced, the keepers of the tradition who have internalized the rules of the design idiom would intuitively recognize the direction of a new example such as those in Figure 5, lines four and five.

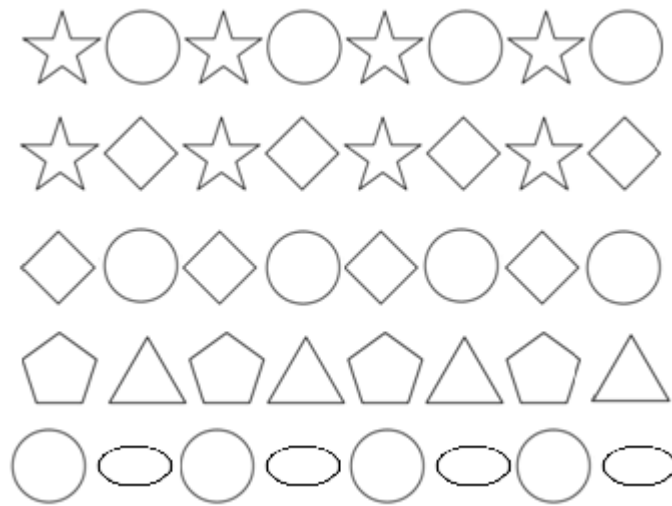


Figure 5, line one: A pattern of stars and circles, representing alternating pointy and round figures in a decorative pattern. (Let's call this order of elements 'forward'.) Line two: A 'forward' pattern incorporating the new rhombus design. Line three: Another forward pattern where the rhombus (previously relatively round) now serves the role of relatively pointy. Line four: A 'reverse' pattern with two new elements. Line five: Another novel "reverse" pattern in keeping with "tradition."

This analogy is limited in its power to portray the workings of music, which must unfold over time and harmonize vertically. Nonetheless, it may help illustrate the clave principle which governs the vertical interactions of simultaneous patterns as well as the horizontal unfolding of the rhythms over time.

A.2.4 The Clave Prototypes: Fundamental Rhythms for Clave Direction

Most studies on the role of clave in Afro-Latin musics limit the exposition of clave to a few fundamental patterns, and neglect to elaborate the mechanism by which *all possible rhythmic sequences* are perceived to be in or out of concord (harmony).

Rather than rely solely on these standard patterns (which can lack a clear link to the arrangement of relative offbeatness in other rhythms), the following examples use more revealing rhythms (*partido-alto*, *gã*, and *son montuno*, Figure 6), where more and less offbeat sections are more apparent than in the *son*, *rumba*, and *bossa* claves¹¹³. They are prototypes that embody direction more clearly than the clave-proper.

The figure displays two musical examples. The top example is a 3-2 rhythm, and the bottom example is a 2-3 rhythm. Each example consists of a musical staff with a treble clef and a key signature of one flat (B-flat), followed by a rhythmic grid. The grids use circles (O) to indicate onbeats and triangles (Δ) to indicate offbeats.

Top Example (3-2 Rhythm):

1	e	&	a	2	e	&	a	3	e	&	a	4	e	&	a
	Δ		Δ			O				O			Δ		Δ

Bottom Example (2-3 Rhythm):

1	e	&	a	2	e	&	a	3	e	&	a	4	e	&	a
		O			Δ		Δ		Δ		Δ			O	

Figure 6: The 3-2 (top) and 2-3 (bottom) characteristic rhythms for samba, with circles indicating onbeats, and triangles indicating offbeats. The 3-2 pattern is what Gomes calls “main rhythmic cell,” and what is meant by *partido-alto* clave direction here is the same idea [161, p. 30].

¹¹³ Just as the English horn is neither a horn nor English (Kerman 1987, 20), perhaps the bossa clave is neither “bossa” nor “clave.” It is not a Cuban pattern, so the word clave is controversial. Furthermore, it is not specific to bossa nova. The typical rhythmic pattern for bossa guitar is the 2-3 form of *partido-alto*. The so-called “bossa” clave is more typically played by the repiques in samba-reggae, by the *caixas* in samba *batucada*, and on *atabaques* in older *carnaval* samba songs.

A.2.4.1 *Partido-Alto*: “Balança” is the essence of *samba carioca*

Though not called a *clave*, *partido-alto* serves the same purpose as *clave* and does so in a more informative manner for mainstream *samba* (Figure 7), so much so that being in *clave* in *samba* is perhaps best called “being in *partido-alto*.” The patterns of Figure 7 are 3-2 because the four on-beats (circles) are on the inside, while the three offbeats (triangles) are on the outside. *Balança* means *swing, as in a playground*. The contour of offbeatness resembles the path followed by a swing: high at the ends, and low in the middle. Furthermore, the greater degree of match with the 3-2 *bossa* *clave* is apparent in the TUBS notation.



	1	e	&	a	2	e	&	a	3	e	&	a	4	e	&	a	
top		H	L	L	L		L		L	H	H	L	L		L		
bottom		Δ			O		O		O			Δ		Δ			
3-2	X			X			X			X			X				
2-3			X			X			X			X			X		

Figure 7: The 3-2 *partido-alto* for *cuíca* and *agogô* (top), and stripped down to its essence (bottom). (The pitches are approximate and can vary with *cuíca* intonation and the *agogô* manufacturer’s accuracy.) Bossa *clave* in 3-2 and 2-3 are added to the TUBS notation. Mismatches with 2-3 *clave* are on the *e* of *two* and the *and* of *four*. The only significant mismatch with the 3-2 *clave* is on the *one*, which is addressed in endnote *ix*.

A.2.4.2 *Son Montuno*: The harmonic backbone of salsa

The *son montuno* pattern, represented here in the most common of its many forms, is the ubiquitous piano riff heard in salsa music. A rhythmic approximation of note attacks, stripping away the chords that would normally be played, is shown in Figure 8.

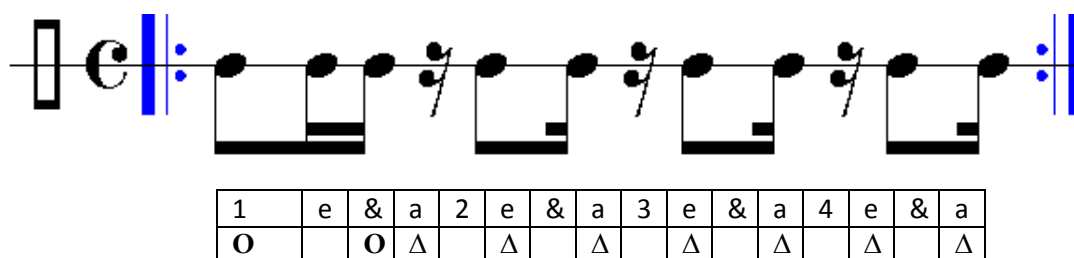


Figure 8: The 2-3 *son montuno* of salsa piano stripped of pitch: The circles represent the only two onbeats (in the outer part of the phrase) and the triangles represent the offbeats which appear in both parts, but only dominate the inner part.

Starting off with two onbeats gives a strong indication that this rhythm will be in the 2-3 clave direction. The introduction of the first offbeat on the *a* of *two*, along with the maintenance of an eighth-note inter-onset interval until the end of the pattern ensures that the inner part of the phrase is exclusively offbeat. This dominance of offbeatness does not let up until the sixteenth-note interval between the very last note and the downbeat at the start of the next repetition. The continuation of offbeatness through the outer part of the pattern at the end can raise the question of whether it is valid to call this pattern 2-3. Again, standard musical practice has this pattern accompanying 2-3 *son* clave in a broad range of salsa pieces from throughout the Americas, and the *relativity* of offbeatness in the wide-sense clave method reconciles this issue decisively: The inner part has higher offbeatness than the outer part. In fact, the only cases of onbeat onsets occur in the outer part, cementing this pattern as 2-3.

The two rhythms depicted and analyzed in this section are of central importance in two prominent examples of Afro-Latin music: samba and salsa.

A.2.5 Wide-Sense Clave Consciousness

A.2.5.1 Implied Clave: Beyond the clave-proper

Thinking of *clave* as a mere collection of traditional patterns, and not the meaning conveyed by these patterns, downplays the cognitive role of clave “within which all the other instruments and voices must fit” (Spiro and Ryan 2006, 14). Almost any pattern in a given clave direction can imply clave sense through the distribution of accents in a percussive or melodic line¹¹⁴. This function of holding the melodic and percussive parts together makes clave an organizing principle worthy of investigation. Requiring more than the detection of specific patterns, it is rather a matter of recognizing implied relationships between patterns. A musician who understands clave can identify the implied direction in any pattern, and one who *feels* clave can do so immediately.

In the earlier stages of developing clave consciousness, if clave-as-pattern is taken literally, the subtleties of wide-sense clave can remain hidden. As discussed above, the most common explanation for the organizational role of clave (involving the first and second halves of the phrase) may fall short of elucidating the workings of clave in many of its cultural manifestations. As a result, students or analysts of clave who follow the standard approach or use the clave-proper solely as a template may find themselves having to make sense of certain musical juxtapositions that diverge from the simple explanation by inventing special cases to explain the perceived inconsistencies (see section 6.4). One may ask why this is a problem—musicians are capable of handling exceptions to rules. However, this is a weakness for a theory of music. As in all phenomena, given two explanations that work in practice, the simpler one that is more consistent (requiring fewer exceptions, if any) is superior. This is the scientifically fundamental principle of parsimony, popularly known as Occam’s razor.

A.2.5.2 Wide-Sense Clave: Beyond percussion

Identifying clave direction is a challenging task for anyone who has not been raised with clave-based music, and occasionally, even for those who have. Furthermore, those who have been culturally trained in clave often find it difficult to explain *why* clave works the way it does. It is not uncommon to hear such explanations as “It’s in the

¹¹⁴ The classification stage of the present research suggests that only on the order of ten percent of all possible rhythmic patterns will fail to establish a sense of clave direction in any musical context.

blood.” However, given our current understanding of the complex interaction of nature and nurture, it is safe to say instead that clave appears to be a rational principle that can be described in a consistent quantitative manner.

This clave sense is not solely percussive, but manifests itself in salsa piano, reggae vocals, bossa nova guitar, *axé* horn arrangements, and in many other ways. It is a rhythmic type of harmony¹¹⁵, exemplified by patterns beyond the typical clave-proper.

This aural tradition of *temporal harmony* is similar to, yet also different from the concept of *tonal harmony* richly cultivated in the European idiom. Clave addresses both static and dynamic relationships (just as tonal harmony does), but these relationships are primarily of note onsets and durations, not pitches. Sublette calls this principle “a key concept of West African music” and a “fundamental structural principle,” and adds that “just as the hardwood claves (the wooden pegs called *clavijas*) once held ships together, when they are clicked together as instruments, the rhythm they play holds the melody line and percussion parts together” (2004, 36). This function of holding the melodic and percussive parts together (*in terms of their timing*) makes clave a regulating principle.

A.2.5.3 Clave Consciousness: Functioning within the idiom

Clave direction constitutes a relationship between offbeatness and time. In common terminology, clave is said to be in “3-2” or “2-3” direction, also called “forward” and “reverse,” respectively. Virtually any rhythm in a duple or quadruple simple meter can be associated with one of these directions. Such association indicates a degree of match with the implied clave.

In practice, the classification of arbitrary rhythm patterns into 3-2 or 2-3 can be a bewildering task. We have observed above that clave gives an indication of relative offbeatness, as opposed to an absolute metric of such. It follows from this observation that the degree of offbeatness in either portion of the clave phrase is meaningful *only with respect to the rest of the phrase*.

As mentioned above, musicians who are clave-conscious do not necessarily need to hear the clave-proper. Clave direction is implied even when the clave is not explicitly voiced. Musicians are *informed* by the clave, and play *around* it, and it is understood that there are “correct” and “incorrect” ways of doing this. In his piano book, Herder calls clave a “demanding and inflexible rule” and refers to it as “the law of clave” (1972, 5).

¹¹⁵ For a similar interpretation of the term ‘harmony’, see “harmonic layers of time” (Leake 2011, 26).

One can, however, choose to break with convention, especially if one knows the rules thoroughly or is functioning outside the idiom. When executed and resolved skillfully, crossing clave is a way to add “spice” to the music, as it is very common in *third-surdo* improvisation in *samba batucada*. Outside of such mindful artistic use, however, crossing clave is frowned upon. Not only for drummers, but for singers, horn players and others, it is not unusual to be asked to leave a session if one crosses clave repeatedly—when it is not called for in the arrangement. (Naturally, this has no bearing on rock or jazz performance as long as a strong link to Afro-Latin tradition is not claimed.)

Like Herder, Mauleón-Santana, author of some of the clearest explanations of clave and its function, differentiates the *instrument*, the *pattern*, and the *concept* of clave (2005, 1). She argues that clave is a rhythmic, melodic, and harmonic rule (2005, 7–9), “albeit a very vague one” (2005, 9). She stresses that clave is omnipresent even when it is not explicitly played (2005, 16), and adds that it is subject to interpretation as well as artistically justified violation (2005, 8–9), as described above.

While in complete agreement with this assessment, the present article is put forth in order to help remove the ambiguity in the role of clave, and to show how the direction of clave can be established or maintained by a variety of patterns.

A.2.5.4 The Relative Nature of Clave: subtlety of an art form

An important step in uncovering the mystery of clave is recognizing that the *complete* clave pattern establishes clave direction. The extent of offbeatness in one section of a phrase is useless for clave direction without knowing the rest of the pattern.

The crux of understanding wide-sense clave is *relative offbeatness* which is in direct contrast to the way clave is typically taught: with two discrete halves, one of which is clearly, absolutely syncopated, and one of which is clearly, absolutely straight.

Unlike in the rigid standard approach, the two sides of clave are not discrete elements. They overlap, sharing inter-onset intervals between them. In light of the greater traditional corpus, and not just a few basic patterns, we abandon the assignment of *absolute* rhythmic tension and release to the first and second halves. The offbeat side of 3-2 clave is *not* the first half; it is a fuzzy temporal region surrounding the first downbeat (the outer part). The onbeat side is a similar zone surrounding the third downbeat (the inner part). Rather than crisp boxes, we have clouds overlapping across bar lines.

Consider the *repique* rhythm of Figure 9, center. In authentic practice, this rhythm is matched with 3-2 patterns in *samba-reggae* and *samba de roda*. In fact, since the *repique* is free to switch from one ostinato to another, this rhythm is *interchangeable* with any of the 3-2 claves. Now note that the latter halves of all three rhythms in Figure 9 are *identical*, but the difference in implied clave direction created by the missing downbeat (shifted onset) in the middle pattern is obvious to the acculturated.

It is not at all uncommon for the clave direction of a pattern to come down to just one semiquaver, as this example shows. In fact, the three patterns belong to three different clave classes. This presents a challenge to group-counting and template-matching for identifying clave direction: How can identical rhythmic cells (the latter halves of Figure 9) be associated with “threeness” in one case, “twoness” in another, and be associated with *neither* direction (lack of clave) in a third case? Should such an exact match with the latter portion of the 2-3 *son* clave indicate that the middle pattern is 2-3, regardless of what the rest of it sounds like? The standard approach to clave may suggest so, but traditional practice suggests otherwise. The analysis presented in this article solves this dilemma decisively.

The rhythmic cell in question is the ubiquitous pattern called *tresillo* or *habanera* in Cuba (Figure 9, bottom). Brazilian audiences and musicians also routinely clap this pattern over *both* halves of any rhythmic cycle. In addition, in the modern Matanzas style of *rumba guaguancó*, the superimposition of the *tumba* and the wooden *claves* gives that same resultant *tresillo* pattern across both sides of *rumba* clave. Hence, one might think that the *tresillo* rhythm is clave-neutral. (It is.)

If so, how is the 3-2 *son* clave not neutral? This is because of relative offbeatness: The *tresillo* call of *son* clave is *more* offbeat than the two-note response, but *less* offbeat than the beginning of the *repique* rhythm shown above. The *tresillo* rhythmic cell, then, is capable of playing the more-offbeat or the less-offbeat role depending on how the rest of the rhythmic cycle is populated. Hence, no one portion of the cycle can be conceived as constituting absolute tension (offbeatness) or resolution. The complete pattern must be considered. The degree of offbeatness in part of a phrase is meaningful only with respect to the rest of that phrase. This idea is one of the keys the solution of the clave riddle.

A.2.6 Relative Nature of Clave Direction in Other Afro-American

Rhythms

The analysis is relevant to the musics of many other cultures in the African Diaspora. In this section, the wider geographical applicability of the proposed analysis is demonstrated using examples from Haiti, Belize, Brazil and the United States. In the process, a key problem in clave direction is resolved in Section 6.4.

The figure displays three musical staves, each with a corresponding 16-beat grid below it. The staves are marked with a 'C' time signature and a blue double bar line. The grids use 'e' for eighth notes, '&' for eighth rests, 'a' for quarter notes, 'O' for half notes, and 'X' for half rests.

Top Staff (2-3 Cuban son clave):

1	e	&	a	2	e	&	a	3	e	&	a	4	e	&	a
		O		O				X			X			O	

Center Staff (3-2 repique ride):

1	e	&	a	2	e	&	a	3	e	&	a	4	e	&	a
	X		X			O		O			O			X	

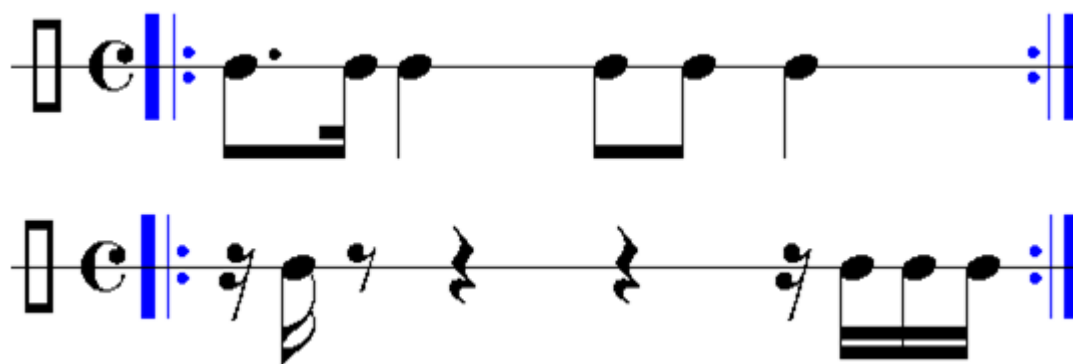
Bottom Staff (clave-neutral tresillo):

1	e	&	a	2	e	&	a	3	e	&	a	4	e	&	a
X			X			X		X			X			X	

Figure 9: The 2-3 Cuban *son* clave (top), a 3-2 *repique* ride typical of *samba-reggae* (center), and the clave-neutral *tresillo* (bottom—a Cuban name for a rhythm found in Haitian *konpa*, Jamaican dancehall, Brazilian *pagode* and *xaxado*, as well as much of Balkan and Roma music).

A.2.6.1 Haitian *Konpa*: A popular Afro-Caribbean form

A selection of instrumental parts that can occur simultaneously in *konpa* appear in Figure 10. The cowbell part is reminiscent of the 3-2 *son* clave, and plays a similar role, reflected in the accompanying hand-drum part which establishes the outer offbeatness that is characteristic of 3-2 clave direction. Among the many variations of this pattern are versions that place open tones on the *e* of *two* and *three*, in direct correspondence with examples shown above.



1	e	&	a	2	e	&	a	3	e	&	a	4	e	&	a
X			X	X				X		X		X			
	T												T	T	T

Figure 10: Common *Konpa* patterns in 3-2, cowbell (top) and tanbou/conga (bottom; 'T' represents an open tone).

A.2.6.2 Funk: Clave in African-American rhythm

In the classic funk beat of Figure 11, top, the prominent “kick” accent on the sixteenth note preceding the omitted downbeat of *three* places this pattern in 2-3 direction.

Another popular funk beat (bottom) is seemingly quite similar to the first example, but is actually in the opposite clave direction. The kick pattern here is the 3-2 *bossa* clave, with the exception that the second note—called *bombo* in Cuba—is missing, leaving only one offbeat in the whole pattern (the *e* of the *four*). The 3-2 feel, based on the presence of this sole offbeat, is reinforced by the “kicks” on the *e*'s of *two* and *three* (onbeats), just as in the *konpa* hand-drum variation and the *partido-alto*.



1	e	&	a	2	e	&	a	3	e	&	a	4	e	&	a
K				S				K				K			

1	e	&	a	2	e	&	a	3	e	&	a	4	e	&	a
K				S				K				S	K		

Figure 11, top: 2-3 funk; bottom: 3-2 funk. *K* stands for “kick” (bass drum); *S* for “snare.”

A.2.6.3 Clave in Belizean Carnival Music

The top line in Figure 12 is a clave pattern from Belize. It functions in the same way as the Cuban “3-2” *son* clave, and it is so close to the latter in sound that this relationship is evident. Yet, the Belizean pattern contains 2 notes in its so-called “3-side” and three notes in its “2-side” (according to the discarded standard method). Clearly, the meaning of clave is not captured by the standard method of counting numbers of onsets in the first and second halves of a phrase.

1	e	&	a	2	e	&	a	3	e	&	a	4	e	&	a
X			X					X		X		X			
X			X			X				X		X			

Figure 12, top: “3-2” clave for Belizean carnival music; bottom: the usual “3-2” *son* clave.

A.2.6.4 An Exception? (*Surdo de Tercera* in Samba *Batucada*)

The *third surdo* as played by the samba schools Beija-Flor, Imperatriz Leopoldinense, Caprichosos de Pilares and Mocidade Independente de Padre Miguel is shown in Figure 13 (Gonçalves and Costa 2000, 48, 49, 52, 54). When using the standard method, this would appear to be clearly 2-3, and is frequently cited thus to argue the absence of the clave concept in Brazilian music. After all, there are two on-beat attacks in the first half, and two offbeat ones in the second half. However, this is a

well-known 3-2 pattern, in that it is played when 3-2 bossa clave is played on the *caixas*, and because native *sambistas* will consider this pattern crossed when the music is in 2-3.

The *surdo de terceira* is a direct descendent of the *partido-alto* (Figure 13, bottom), generated by omitting the first and third quarters of the *partido-alto* cycle. (Such deep connections between the patterns of various instruments is the source of the present approach.) Indeed, the offbeat section of the rhythm falls on the “outer part” of the phrase, and the onbeat section on the “inner part” of the phrase (see Figure 13).

1	e	&	a	2	e	&	a	3	e	&	a	4	e	&	a
				L		L						L	L		L

1	e	&	a	2	e	&	a	3	e	&	a	4	e	&	a
	H	L	L	L		L		L		H	H	L	L		L

Figure 13, top: The *surdo de terceira*. Bottom: The 3-2 *partido-alto* for *cuíca* and *agogô*. Note that playing the *partido-alto* omitting the first and third crotchet's worth of onsets results in the *terceira*.

A.2.6.5 Subtle Examples from *Batucada* and *Samba de Roda*

Two final examples are included to further demonstrate *relative* offbeatness, and to emphasize the close connection between the Afro-Latin rhythm patterns of different cultures. The top pattern shows only one onset at the very end of the cycle breaking the eighth-note accent pattern. An examination of what the other drums in the Tijuca and Viradouro samba schools play will reveal this pattern to be in the 3-2 direction. The

standard approach would place this rhythm in the 2-3 category because the only accent of interest is in the second half of the phrase. The present approach, however, elegantly explains traditional practice when the standard approach fails.

The figure consists of two musical examples, each with a staff and a corresponding table of accents.

Top Example: caixa pattern

The staff shows a sequence of notes: a half note, followed by four eighth notes, and a dotted quarter note. The notes are grouped into four measures of two eighth notes each.

1	e	&	a	2	e	&	a	3	e	&	a	4	e	&	a
X		X		X		X		X		X		X		X	

Bottom Example: gã (bell) pattern

The staff shows a sequence of notes: a half note, followed by four eighth notes, and a dotted quarter note. The notes are grouped into four measures of two eighth notes each.

1	e	&	a	2	e	&	a	3	e	&	a	4	e	&	a
X		X	X	X		X		X		X		X	X		

Figure 14, top: The accents for the *caixa* pattern of samba schools Unidos Da Tijuca and Unidos Do Viradouro (Gonçalves and Costa 2000, 59, 61). Since the only note breaking the steady pattern of *on*beats is the last one, the pattern is in 3-2. This is an excellent authentic example of clave direction as *relative* offbeatness.

Bottom: The 2-3 gã (bell) in *samba de roda*, an example of how offbeatness can threaten to destroy the underlying pulse and establish a new one (Merriam 1959, 16), and how the re-establishment of the original *tactus* generates the clave feel.

The bottom pattern from the gã (bell) in traditional *samba de roda* is almost exactly the same as the *son montuno* of salsa (Figure 8), except that it resolves back to *on*beats an eighth note sooner, establishing even higher *on*beatness in the outer part of the pattern to contrast with the relentless offbeatness of the inner part. Hence, this pattern is almost the same as the 2-3 *son montuno*, but even more clearly 2-3. Stripping the notes of their pitches thus reveals the cross-cultural links among Afro-Latin forms.

A.2.7 Conclusion

Clave need not be as elusive (Mauleón-Santana 2005, v), vague (Ibid., 9), unforgiving (Herder 1972, 5, 12), or seemingly replete with inconsistencies as many authors have suggested. We hope that with this exposition, the path toward a consistent, parsimonious theory of clave in particular and Afro-Latin rhythm in general has been initiated, and process of internalizing clave (for listeners, as well as performers, teachers, students and musicologists) may proceed more smoothly than it might otherwise.

The main points made in this article are that 1) the absolute onset count should be abandoned early in the process of understanding clave; 2) when sixteenth notes are the smallest subdivision, upbeats and downbeats are *on*beats, as far as clave sense is concerned; 3) the inner/outer demarcation suggested in Figure 2 should be preferred over the standard approach of the first and second halves of the phrase; and 4) it is the *relative* offbeatness of sections that determine the sense of clave, not counting onsets or template-matching against well-known patterns.

Armed with these introductory notions, the interested reader can broaden their understanding of the manifestations of clave in other cultural idioms of the African Diaspora. While the more specific ideas in this article are primarily discussed through Cuban, Brazilian and a few other examples, the principles of the wide-sense framework may be applied to the musical systems of many Afro-Latin styles, as they were applied briefly to Haitian, Belizean, Ghanaian (end note *iv*) and North American styles here.

There is more to clave than discussed in this article. For example, clave has timing and phrasing functions in addition to what is explored here. Furthermore, the existence of *clave-ambiguous* patterns (in terms of *clave function*) have been identified by the author, and listening trials with leading experts suggest that clave relationships depend on a number of schema-type rules based on short rhythmic cells in key positions. The interaction of such rules as applied to vague patterns goes beyond the basics of clave discussed above. These guidelines can easily conflict when their characteristic rhythmic cells coexist in one pattern, leading to a case of context-dependence in clave-direction. An example of this can be found in the *cáscara matancera* (for example, Mauleón 1993, 76–78) whose overall pattern agrees with the analysis offered here, but focusing on sixteenth-note components of each beat reveals that the single most offbeat section is on the inside in 3-2 *cáscara*. Hence, detailed implementation can vary from culture to culture even when the overall ebb and flow of offbeatness conforms across cultures.

Research is also required to extend this system to ternary patterns, and the conversion appears non-trivial. Future work should also incorporate note releases into a more general theory of clave for instruments capable of sustained tones.

Nonetheless, it is germane to point out that the concepts of offbeatness and relativeness, and the “inside/outside” perspective (rather than first half/second half) seem to explain the culturally accepted clave directions of many more patterns than any previous analysis or theory of clave.

References for Appendix A

- Anku, W., "Towards a Cross-Cultural Theory of Rhythm in African Drumming," In *Intercultural Music*, edited by C. T. Kimberlin and Akin Euba, 167-202. Bayreuth, Germany: Breitingen, 1995.
- Baroni, M., "The Concept of Musical Grammar," *Music Analysis*, Vol. 2, No. 2, pp. 175–208, 1983.
- Collins, J., "African Musical Symbolism in Contemporary Perspective: Roots, Rhythms and Relativity," Berlin, Germany: Pro BUSINESS GmbH, 2004.
- Cook, G., *Teaching Percussion*, New York, NY: Schirmer Books, A Division of Macmillan, Inc., 1988.
- Cruz, T., Moore, K., Gerald, M., and Fiol, O., *The Tomás Cruz Conga Method: Conga Technique as Taught in Cuba - Intermediate: Essential Cuban Conga Rhythms*, (3 vols.) Vol. 2. Pacific, MO: Mel Bay Publications, Inc., 2004.
- Ekwueme, L. E. N., "Concepts of African Musical Theory," *Journal of Black Studies* Vol. 5, No. 1, 1974, pp. 35–64.
- Gomes, S., *New Ways of Brazilian Drumming -Nuevos Caminos De La Bateria Brasileña*, Tübingen, Germany: advance music, 2007.
- Gonçalves, G., and Costa, O., *Aprendendo a Tocar O Batuque Carioca: As Baterias Das Escolas De Samba Do Rio De Janeiro*, Brazil: Groove, 2000.
- Herder, R., *Spanish Montunos – Ensemble Studies in Rhythmic Precision and Pitch Security: Solo Book of Montuno Improvisation for Concert Band, Stage Band, Orchestra or Any Instrumental Combination*, New York, NY: Continuo Music Press, 1972.
- Kauffman, R., "African Rhythm: A Reassessment," *Ethnomusicology* Vol. 24, No. 3, 1980, pp. 393–415.
- Kerman, J., *Listen (Brief Ed)*, New York, NY: Worth Publishers, Inc., 1987, p. 20.
- Koetting, J., "Analysis and Notation of West African Drum Ensemble Music," *Selected Reports in Ethnomusicology* Vol. 1, No. 3, 1970, pp. 115–46.
- Koetting, J., and Knight, R., "What Do We Know About African Rhythm?" *Ethnomusicology* Vol. 30, No. 1, 1986, pp. 58–63.

Kwabena Nketia, J. H., *African Music in Ghana*. Chicago, IL: Northwestern University Press, 1970.

Ladzekpo, C. K., *Rhythmic Principles (Foundation Courses in African Music)*, 1985.
Available: www.cnmat.berkeley.edu/~ladzekpo/PrinciplesText.html

Leake, J., "Harmonic Time: Multidimensional Awareness of Polyrhythms, Polytempos and Polyfeels," *Percussive Notes: the Journal of the Percussive Arts Society* Vol. 49, No. 1, 2011, p. 26.

Lehmann, B., *The Syntax of 'Clave' – Perception and Analysis of Meter in Cuban and African Music*, thesis, Tufts University, 2002.

Locke, D., *Drum Gabu—The Rhythms of West African Drumming*, Crown Point, IN: White Cliffs Media Company, 1987.

London, J., *Hearing In Time: Psychological Aspects of Musical Meter*, New York, NY: Oxford University Press, 2004.

Machado, H. "Foca", Muñoz, W., and Sadi, J., *El Toque De Candombe: Ritmo Tradicional Afro-Uruguayo / Traditional Afro-Uruguayan Rhythm*, Pacific, MO: Mel Bay Publications, Inc., 2002.

Malabe, F., and Wiener, B., *Afro-Cuban Rhythms for Drumset*, Van Nuys, CA: Alfred Publishing Co., Inc., 1990.

Mauleón, R., *Salsa Guidebook for Piano & Ensemble*, Petaluma, CA: Sher Music Co., 1993.

Mauleón-Santana, R., *101 Montunos*. Petaluma, CA: Sher Music Co., 2005.

Merriam, A. P., "Characteristics of African Music," *Journal of the International Folk Music Council*, No. 11, 1959, p. 13–19.

Novotney, E. D., *The 3:2 Relationship as the Foundation of Timelines in West African Musics*, dissertation, University of Illinois at Urbana-Champaign, 1998.

Rodriguez, A., *Traditional Afro-Cuban Concepts in Contemporary Music (Ethnic Percussion Series)*, Pacific, MO: Mel Bay Publications, Inc., 2003.

Rosauro, N., *The ABCs of Brazilian Percussion: The Easiest Way to Teach Yourself How to Play the Essential Brazilian Percussion Instruments*, Brazil: Propercussão Brasil, 2004.

Soebbing, H. W., "Latin American Instruments and Rhythms," *The Instrumentalist* (Dec. 1977) in *Percussion Anthology—A Compendium of Percussion Articles from The Instrumentalist*

Fourth Ed., including articles from 1946 through 1987, Northfield, IL: The Instrumentalist Co., 1988.

Spiro, M., and Ryan, J., *The Conga Drummer's Guidebook*, Petaluma, CA: Sher Music Co., 2006.

Sublette, N., *Cuba and Its Music: From the First Drums to the Mambo*, Chicago, IL: Chicago Review Press, 2004.

Tobias, D. A., "Latin-American Percussion," (1965) *The Instrumentalist* (Nov. 1965) in *Percussion Anthology – A Compendium of Percussion Articles from The Instrumentalist Fourth Ed., including articles from 1946 through 1987*, Northfield, IL: The Instrumentalist Co., 1988.

Toussaint, G. T., "A Mathematical Analysis of African, Brazilian, and Cuban Clave Rhythms," Paper presented at the *BRIDGES: Mathematical Connections in Art, Music and Science*, Towson University, Towson, MD, 2002.

Toussaint, G. T., "Mathematical Features for Recognizing Preference in Sub-Saharan African Traditional Rhythm Timelines," *Lecture Notes in Computer Science: 3686*, 2005, pp. 18–27.

Vurkaç, M., "Clave-direction analysis: A new arena for educational and creative applications of music technology," *The Journal of Music, Technology and Education*, Vol. 4, No. 1, 2011, pp. 27–46.

Washburne, C., "The Clave of Jazz: A Caribbean Contribution to the Rhythmic Foundation of an African-American Music," *Black Music Research Journal*, Vol. 17, No. 1, 1997, p. 66.

APPENDIX B: Information Theory

This appendix tackles Information Theory with a tutorial-type approach, with a number of intuitive examples and the clarification of key distinctions among certain information-theoretic quantities.

B.1 Uncertainty and Information

Based on Probability and Statistics, and primarily developed as a theory for Communications and Signal Processing applications [1], Information Theory has come to play an important role in a number of scientific, social, and technical fields, from Psychology and Sociology to Computer Science and Economics [2–5, respectively].

It is essential to first clarify that *Information*¹¹⁶ in the information-theoretic sense is separate from *meaning*¹¹⁷, just as the term *statistically significant* does not necessarily imply sociological or medical significance. Weaver expresses the distinction in the following manner: “This word information in communication theory relates not so much to what you *do* say, as to what you *could* say. That is, information is a measure of one’s freedom of choice when one selects a message” [1]. Similarly, Miller explains Information as a measure of organization in that “a perfectly organized system is completely predictable and its behavior provides no [new] information at all.” [2, p. 123] He goes on to say that “the more disorganized and unpredictable a system is, the more information you can get by watching it.” (Ibid.) In this sense, it is important to note that any scientific “measurement is communication between nature and scientist” [6].

¹¹⁶ Terms such as ‘information’ ‘constraint’ and ‘uncertainty’ are selectively capitalized throughout this document to distinguish strict information-theoretic use from everyday use. Similarly, terms like ‘neural networks’ and ‘genetic algorithms’ are capitalized when referring to the field of study and not when referring to individual instantiations of networks or algorithms.

This practice should not be considered excessively esoteric—compare this with the defensible distinction made among “MUSIC, Music, and music” by Elliott in his book on music education: “MUSIC is a diverse human practice consisting in many different musical practices or Musics. Each and every musical practice (or Music) involves the two corresponding and mutually reinforcing activities of music making and music listening. [...] The word music (lowercase) refers to the audible sound events, works, or listenables that eventuate from the efforts of musical practitioners in the contexts of particular practices.” [7]

¹¹⁷ ‘Meaning’ can be thought of as the co-existence of high Information and high Constraint, which is not the same as purely high Information. [6]

This ‘amount of information’ one can get from observing or measuring is related to the amount of uncertainty that the system possesses. Uncertainty in Information Theory has to do with the potential variety of outcomes, outputs, or actions that may result from interactions with a system, and with its *potential* to provide Information to an observer.

Equation (22) is Shannon’s expression for Entropy¹¹⁸. The meaning of this equation can be understood intuitively if one considers a loaded coin and a fair coin undergoing a large number of tosses.

$$H(x) = -\sum_{j=1}^n p(x_j) \log [p(x_j)] \quad (22)$$

In (22), we see that Uncertainty, related to the amount of surprise in finding out that a random variable x takes on the value x_j (which is the quantity of the logarithm), is the *average surprise* over all trials.

Let the former be loaded such that the probability of heads, $P(H)$, is 0.01 and the probability of tails, $P(T)$, is 0.99. Although getting a heads is highly surprising, the probability of getting tails is so high (the event is so ordinary) that average surprise over many trials is low: $H(X) = -(0.01 \log_2(0.01) + 0.99 \log_2(0.99)) = 8.08 \times 10^{-2}$.

For the fair coin, however, $H(X) = -(0.5 \log_2(0.5) + 0.5 \log_2(0.5)) = 1$. On average, we are more surprised with the fair coin.

The amount of Uncertainty is a property of a system as a whole, or of a state that a system may be in. Individual messages (such as hypotheses, models, networks, patterns, training examples, categories, measurements, etc.) on the other hand, carry Information, the transmission and reception of which may change the amount of Uncertainty (disorganization and potential for information) in the system being observed. For equally likely outcomes, Uncertainty is the number of yes/no questions it would take to get the (correct) answer. (Hence, “Uncertainty” is a numerical value.)

Hence, it is necessary to quantify the relationship between Uncertainty (disorganization, or Entropy¹¹⁹), and Information (Constraint, or loss of Entropy).

¹¹⁸ One may refer to Gleick’s *The Information* for the interesting history behind Shannon’s discovery [8], and to Schneer [9, for a clear explanation (using mixtures of gasses) of how entropy in thermodynamics is analogous to information-theoretic entropy.]

Frequently in applications of Information Theory, the word ‘information’ is used in reference to both Information and Uncertainty. This common practice has a simple mathematical reason, though it leads to a great deal of confusion. It is necessary to distinguish information in the sense of *potential for information* (closely related to Uncertainty, disorganization, and Entropy), and *information-in-the-everyday-sense* (closely related to organization, or loss of Entropy). In general, any occurrence of the term Information in the preceding paragraphs can be replaced by *information-in-the-everyday-sense*, while occurrences of the term Uncertainty can be replaced by *potential for information*. Common practice is to simply say ‘information’ in both cases. The justification for this overloading is found in the following identity (23).

$$\text{Information} = -\Delta H = -(H_{\text{final}} - H_{\text{initial}}) = H_{\text{initial}} - H_{\text{final}} \quad (23)$$

This is the source of the mistaken equivocation of Information and Uncertainty, where the latter is represented by the variable H , which is also Entropy.

Note that when $H_{\text{final}} = 0$ (that is, when so much information is received that no Entropy remains), numerically speaking, $\text{Information} = H_{\text{initial}}$. In other words, if a system starts out in a state of uncertainty and then all uncertainty is removed by the arrival of some information, the total uncertainty removed is numerically equal to the total information received.

This special case is thought of as a generality when the term ‘information’ is used to mean Uncertainty or Entropy. In fact, Uncertainty is not the same as Information because Information is the *negative* of the *change* in Uncertainty.

When Information and Uncertainty are numerically equal, this special case has led to the common interpretation of Information Theory as if it claims that *Information is Uncertainty*. Misleading as it is, this is a helpful shortcut as long as the underlying condition is understood.

Cover and Thomas help clarify the issue with the following statement: “The entropy of a random variable is a measure of the uncertainty of the random variable; it is a measure of the amount of *information required* on the average to describe the random variable” [10, p. 18; emphasis added]. Here, ‘information required’ is contrasted with ‘information possessed’, or ‘information contained’.

An example involving Uncertainty and the transmission of Information, along with the interplay of Constraint and Entropy would be the case of a lecture. The

¹¹⁹ All uses of the word ‘Entropy’ in this dissertation refer to “Shannon Entropy,” not thermodynamic entropy, the difference being that in Information Theory, closed systems whose entropy monotonically increases are not a necessity.

audience knows that the lecture has a particular topic. This is a form of constraint (lowercase in this example because precise mathematical metrics are not used in this example). If the topic is, for instance, Laplace transforms, the audience knows not to expect any mention of cooking or fashion. The probabilities assigned to those topics are near zero. Very high probabilities are assigned to exponentials, electricity, system response or stability. Nonetheless, no audience member knows the exact words that will be spoken by the lecturer, or even the sequence of ideas that will be dealt with. This is uncertainty. When there exists a balance between uncertainty and constraint, meaning may be transmitted. As each sentence is transmitted by the lecturer, a small amount of information is delivered to the audience, reducing the lecturer-system's uncertainty by that amount. In this example, the word "uncertainty" has been used from the point of view of the audience. The same quantity, from the point view of the lecturer-system (the intersection of lecturer and his/her domain knowledge) would be called entropy because it reflects the complexity of the system that generates the sequence of information-bearing messages.

To derive a mathematically precise definition for this systemic uncertainty, consider Miller's explanation of the *amount of information* via the example of "a child who is told that a piece of candy is under one of 16 boxes" [2, p. 124]. This experimental setup represents a particular degree of uncertainty, albeit not yet quantified.

The more information the child has, the better a decision he or she can make about which box to lift. If one of the 16 boxes is eliminated by such information supplied to the child, the number of choices is only reduced by one. Should this be considered one unit of Information? Would it also be one unit of information had there been a thousand boxes, or only 3? We can see that it is not preferable to adopt such a convention. If another child were presented only two boxes, and given the same additional knowledge that one particular box is undesirable, the effective amount of Information gained by the child facing two boxes would be substantially greater than in the case of the child facing 16 boxes. It seems that the number of alternatives that existed to begin with is as important as the number of alternatives removed from consideration. Therefore, a ratio of (rather than an absolute change in the numbers of) alternatives is preferable as a measure of Information [2, p. 124].

Since the *amount of information* (amount of Constraint) has to do with the extent to which a particular message reduces the number of possible messages deliverable, and the appropriate measure of such involves a ratio, the concept of *amount of information* is understood to mean that: "every time the number of alternatives is reduced to half, one unit of information is gained ... called one 'bit' of information" [2, p. 124]. Note that this information is transferred from the system to the observer by the action of the

message delivered. Information gained by the observer is system Uncertainty reduced, system potential lost, or system Constraint increased.

Constraint also relates to the impact of the term *given* when one speaks of conditional probabilities. Constraint is how much knowing one variable reduces the variety in the other variable. Constraint captured (Information received) is Uncertainty reduced.

In the case of the child with 16 boxes, a message that reduces the number of desirable alternatives to 2 contains one bit more Information than another message that only reduces the number of alternatives to 4. By noting that going from 16 to 2 involves halving the number of alternatives three times, whereas going from 16 to 4 involves halving twice, we have an intuitive explanation for the amount of Information in selecting a message as $\log_2 N$ [2, p. 124].

Hence, $U(x)^{120}$, the amount of information required to describe a random variable x , or the uncertainty in selecting a message x is (24):

$$\log_2 = U(x) \quad (24)$$

where the number 2 represents the dichotomy of ‘Yes’ and ‘No’, and x represents the variety of alternatives. In other words, the logarithm of the number of alternatives is the amount of information. Similarly, the binary choice, raised to the power of the amount of information equals the number of alternatives:

$$2^0 = 1 \text{ (no uncertainty means no alternative)}$$

$$2^{10} = 1024 \text{ (some uncertainty means some alternatives)}$$

Thus, the amount of uncertainty can be thought of as the number of unknowns in a binary system. If the system under consideration is not binary, then uncertainty is proportional to the number of unknowns.

To put it another way, uncertainty measures “the *logical variety* in a descriptive system of categories” [11, p.13] in a discrete-valued, nominal system where the values of

¹²⁰ The letters ‘U’ and ‘H’ are used interchangeably by most theorists and practitioners of Information Theory because the distinction between Uncertainty (U) and Entropy (H) is not shared by all. Both letters are used here to draw attention to yet another instance of equivocation *and* equivoque in Information Theory. Furthermore, the attempt by McGill and Quastler to come up with a standardized nomenclature [13], in the opinion of this author, only made the issue more confusing.

variables are categories. In fact, Uncertainty is also thought of as a measure of diversity, or of uniformity of spread [6] and is thus related to variance (σ^2).

Out of a set of possible messages, X , the amount of information a message x conveys (25) is the difference between the uncertainty in the system *without* (before) knowledge of x and the uncertainty in the system *given* (new) knowledge of x [11, p. 13].

$$I(x \in X) = U(X) - U(x) = \log_2 N_X - \log_2 N_x \quad (25)$$

By the properties of logarithms, we then have (26):

$$\log_2 N_X - \log_2 N_x = -\log_2 \frac{N_x}{N_X} \quad (26)$$

Thus, information turns out to be “a function of the probability of selecting the desired set of alternatives by chance” [11, p.14] under an assumption of equal likelihood; i.e., without taking frequencies of occurrence into account.

B.2 Entropy

Entropy is related to Information in that it is a measure of diversity, but while Uncertainty concerns the *possible* variability of messages, *Entropy* concerns the *actual* (observed) variety: “Part of the definition of entropy, and one reason for calling it a measure of observational variety, is that unobserved possibilities do not enter the measure” [11, p. 16]. In this sense, Uncertainty and Entropy can even be exact opposites. Yet, they are often conflated (see below).

Because of the difference in the consideration for unobserved alternatives, Entropy incorporates frequencies of occurrence, while Information does not [11, p. 15].

With $2^{16} = 65536$ cases treated individually, the Uncertainty is 16 bits. However, if we were to classify these 65536 instances into four categories, $\{c_1, c_2, c_3, c_4\}$, with 25379 members in c_1 , 20313 members in c_2 , 4157 members in c_3 , and 15687^{121} members in c_4 , the Uncertainty reduced by the first cluster is $\log_2 25379 = 14.631$ bits; the Uncertainty reduced by the second cluster is $\log_2 20313 = 14.310$; the Uncertainty reduced by the third cluster is $\log_2 4157 = 12.021$; and the Uncertainty reduced by the fourth cluster is $\log_2 15687 = 13.937$.

¹²¹ These numbers are purely hypotheticalal, selected for the sake of example.

Clearly, it does not make sense to say that we started with an Uncertainty of 16 bits, and reduced it by $(14.631 + 14.310 + 12.021 + 13.937) = 54.899$ bits.

A more reasonable way to calculate Uncertainty reduction is to compare Uncertainty values before and after classification. Prior to classification, there were 65536 separate categories, since each object was its own category. After classification into 4 categories, we can say that the uncertainty has been reduced to $\log_2 4 = 2$. However, this value does not take into account the frequencies of occurrence of the members of the four categories. The ideal measurement of the Uncertainty reduction should be close to 2, but incorporate how the objects are distributed into the four categories.

One expression for Entropy (observed Uncertainty) is given by (27)[11, p. 16]:

$$H(A) = \log_2 n - \sum_{a \in A} \frac{n_a}{n} (\log_2 n_a) = \sum_{a \in A} \frac{n_a}{n} \left(-\log_2 \frac{n_a}{n} \right) \quad (27)$$

Applied to the four-way classification described above, $H(A) = 1.799$. This is an average of the Information content of each category, weighted by the frequency of occurrence of objects in each category. Hence, Entropy is “the average amount of information required to select (predict or identify) observations by categories” [11, p. 16].

Frequencies of occurrence are not the same as probabilities. However, in the limit as n approaches infinity, one can treat the former as the latter. In other words, frequencies of occurrence *are* probabilities when taken asymptotically.

Entropy is maximal when all categories are inhabited by exactly the same number of objects. This includes the case where each category is inhabited by a single example, in which case, Entropy equals Uncertainty. Hence, at least according to the present author’s interpretation of Krippendorff, Uncertainty is information-theoretic variety while Entropy is set-theoretic variety. Both are nominal measures of diversity, or spread, in comparison with variance, which is a numerical measure of spread.

Figure 65 is a depiction of the difference between entropy and variance for abstract cases of continuous and discrete variables: parts a and b have the same entropy but different variance (greater for b). The same goes for parts c and d .

B.3 Transmission, Mutual Information, Information Distance and the Kullback-Leibler Divergence

Four closely related and often-confused measures of Information (hence of Constraint, Uncertainty and Entropy) are Transmission, Mutual Information, Information Distance and the Kullback-Leibler Divergence (also misleadingly called Kullback-Leibler *Distance* [12, p. 8]).

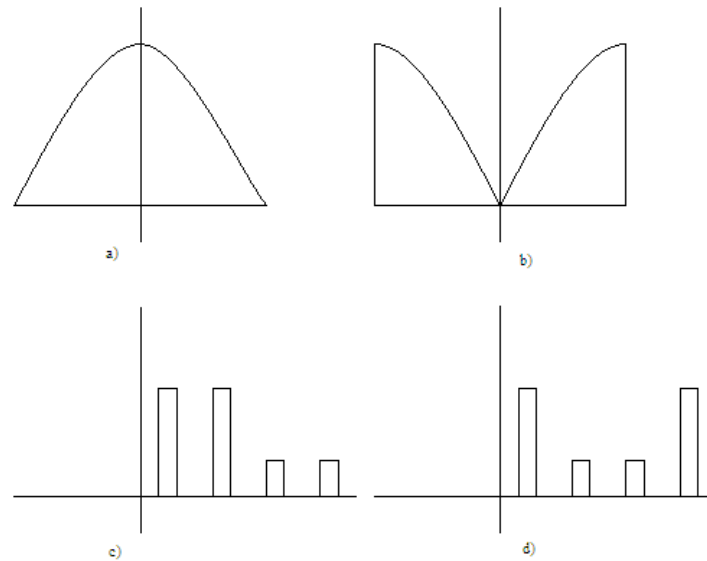


Figure 65: Entropy is the same for *a* and *b* (continuous) and for *c* and *d* (discrete), while variance is greater for *b* and for *d*.

The primary difference between these measures is their direction. Transmission always measures top down (from the data, or maximal constraint), and reflects the error in a model *as compared to the data* on hand. Information Distance, on the other hand, can range from any point to any point (model to model, data to model, model to data, independence to model, data to independence, independence to uniformity, uniformity to model, etc.).

Transmission is the measure of *Constraint lost* in going from the data to independence, or to a model. It is also numerically equal to *Constraint captured* in going from independence (or a model) up to the data. This, in turn, is equivalent to the Uncertainty reduced in going from independence (or a model) to the data, and the

Uncertainty increased in going from the data to independence (or a model, because a model typically has fewer parameters, so fewer degrees of freedom). Transmission always measures from the data down, and reflects the amount of error in a model in relation to the data, where “data” implies maximal Constraint.

Information Distance, on the other hand, concerns arbitrary levels along the lattice of structures, but always points down the lattice because it is a difference of Transmission values.

In Figure 66, the Information Distance represented by the circle is the amount of Constraint modeled in m_i but not captured by m_j given the data. Similarly, the Transmission of the model m_i is the amount of Constraint in the data (m_0) that escapes explanation by m_j (the error in m_j).

Figure 66 shows the relationships between Entropy and Transmission. $H(X:Y)$ is the Uncertainty in the model where X and Y are independent. $H(XY)$ is the Uncertainty in the model XY , which means that an interaction effect exists between the two variables. In the case where those are the only two variables in a system (see Figure 67), the Transmission $T(X:Y)$ is the error in the independence model, which, by definition, fails to capture the interaction between X and Y .

Equation (28) shows the Information Distance from m_i to m_j : One model has captured greater Constraint than the other.

$$I(m_i \rightarrow m_j) = I(m_0 \rightarrow m_j) - I(m_0 \rightarrow m_i) = T(m_j) - T(m_i) \quad (28)$$

With respect to (29), Transmission $T(X:Y)$ is the error in the independence model which fails to capture the interaction between X and Y . $H(X:Y)$ is the Uncertainty in the model where X and Y are independent. $H(XY)$ is the Uncertainty in the model XY , which means that an interaction effect exists between the two variables.

$$T(X : Y) = H(X : Y) - H(XY) = H_{\text{mod}} - H_{\text{data}} = H(X) + H(Y) - H(XY) \quad (29)$$

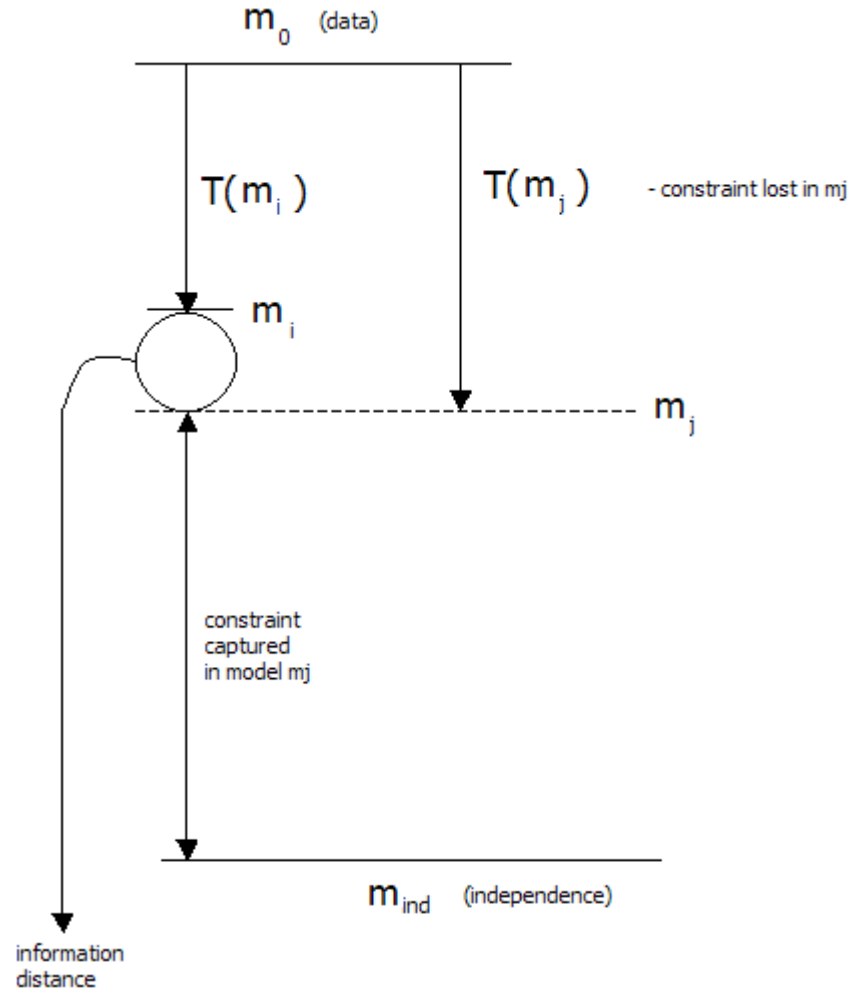


Figure 66: Transmission and Information Distance

The third expression in (29) more clearly indicates this error as a difference of Uncertainties. The final (fourth) expression in (29) and all of (30) refers to the Uncertainty circles in Figure 67.

$$T(X : Y) = H(XY) - H(X | Y) - H(Y | X) \quad (30)$$

Though they may look like it, Uncertainty circles are not Venn diagrams¹²². Each circle's area represents an amount of Uncertainty, or Entropy, and the area of the intersection represents the amount of Transmission. In the traditional Shannon Information Theory sense, a message X is transmitted, and a message Y is received. The higher the fidelity of the channel, the greater the intersection, and thus its Transmission. $H(x|y)$ is the Uncertainty in what message was sent, given that a certain message, Y , was received. This is called *equivocation*. $H(y|x)$ is the Uncertainty in the received message, given that X had been sent. This is called *noise*. The interaction between the variables (messages) is reflected in the system Uncertainty $H(x,y)$.

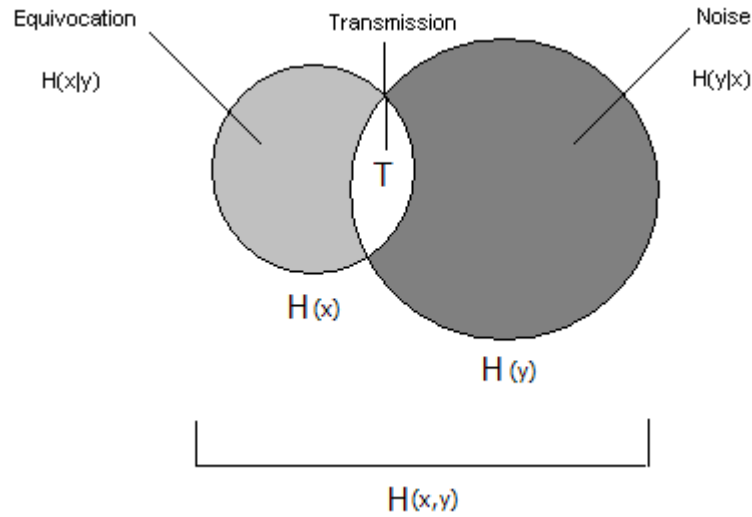


Figure 67: Uncertainty circles depicting Transmission, equivocation and noise in a two-variable system.

¹²² There are two significant differences between Uncertainty circles and Venn diagrams. The intersection of more than two Uncertainty circles can have *negative* area (whereas Venn diagrams cannot). Secondly, disjoint Venn diagrams represent probabilistically *dependent* events, whereas disjoint Uncertainty circles represent *independent* variables.

Mutual Information is a redundant concept that is not additionally informative in the presence of Transmission and Information Distance. It is defined as the decrease in Uncertainty about one variable based on observing another variable [12, p. 7], which is simply Transmission written in a form that is a hybrid of the conventions for Transmission and Information Distance. In (31), Mutual Information is given in terms of Entropies. Note the semicolon notation.

$$I(X;Y) = H(X) - H(X | Y) = H(Y) - H(Y | X) = H(X) + H(Y) - H(XY) = T(X : Y) \quad (31)$$

In (32), we show how Mutual Information, Transmission and Information Distance are equivalent. Information Distance is the most informative in terms of its expressional convention.

$$I(X;Y) = T(X : Y) = I(XW : XY \rightarrow XW : Y) \quad (32)$$

Furthermore, all of these measures are equivalent to the Kullback-Leibler Divergence. By convention, Mutual Information is mostly used to mean the entire Transmission from the data to independence, and the Kullback-Leibler Divergence is Transmission expressed in terms of probability distributions [10, p. 18; 101, p. 8].

References for Appendix B

- [1] Shannon, C. E., and Weaver, W., *The Mathematical Theory of Communication*, Urbana, IL: The University of Illinois Press, 1964.
- [2] Miller, G. A., “What is Information Measurement,” *American Psychologist*, No. 8, 1963, pp. 3–11.
- [3] Galtung, J., *Peace, War and Defense: Essays in Peace Research, Vol. I: Peace, Research, Education, Action*, Copenhagen, Denmark: Christian Ejlertsen, 1975.
- [4] Quinlan, J. R., “Induction of Decision Trees,” *Machine Learning*, No. 1, 1986, pp. 81–106.
- [5] Theil, H., *Statistical Decomposition Analysis*, New York, NY: Elsevier, 1972.
- [6] Zwick, M., (class notes), Discrete Multivariate Modeling, PSU SySc 551/651 January 15, 2008.
- [7] Elliott, D. J., *Music Matters: A New Philosophy of Music Education*, New York/Oxford: Oxford University Press, 1995, pp. 44–45.
- [8] Gleick, J., *The Information: A History, A Theory, A Flood*, New York: Pantheon Books, 2011, pp. 4–8.
- [9] Schneer, C., J., *The Evolution of Physical Science: Major Ideas from Earliest Times to the Present*, Lanham, MA: University Press of America, 1984, pp. 256–259.
- [10] Cover, T. M., and Thomas, J. A., *Elements of Information Theory*, New York, NY: John Wiley & Sons, Inc., 1991, p. 18.
- [11] Krippendorff, K., *Information Theory: Structural Models for Qualitative Data, (Quantitative Applications in the Social Sciences #62)*, Sage, Beverly Hills, 1986.
- [12] Doya, K., Ishii, S., Pouget, A., and Rao, R. P. N., *Bayesian Brain: Probabilistic Approaches to Neural Coding*, Cambridge, MA: The MIT Press, 2007.
- [13] McGill, W. J., and Quastler, H., “Standardized nomenclature: an attempt,” In H. Quastler (Ed.), *Information Theory in Psychology*, Glencoe, IL: Free Press, 1955.

APPENDIX C: Literature Review of Music Information Retrieval and Digital (Audio) Signal Processing

There are eight primary objectives evidenced in the papers collected for the MIR/DSP literature review. They are the identification, detection or tracking of the following:

- Beat (Pulse, Tactus, or Foot-Tapping Rate)
- Tempo
- Meter (or Time Signature)
- Tatum
- Swing (Shuffle, Rubato, or Feel)
- Tempo Modulation (Accelerando/Ritardando)
- Timbre
- Genre

The papers reviewed for this report can be categorized in several ways:

- Fully implemented vs. demonstrative (partially implemented)
- Frequency-domain vs. time-domain
- Engineering vs. psychoacoustical
- Genre-specific vs. no constraints
- Audio data vs. MIDI data

The demonstrative papers rely on human interpretation of plots or charts for the final conclusion on the success rate of an algorithm. The fully implemented papers have the data-analysis and decision-making aspects of rhythm tracking implemented in addition to the processing of music input.

The frequency-domain papers clearly use models of human perception while the time-domain papers typically analyze data in the time domain without help from music theory or psychology.

The engineering papers are concerned with machine perception while the psychological papers examine human perception of rhythm, and help form models of potential machine perception.

Some research targets specific genres or audio that meets certain constraints and requirements, such as constant tempo, non-compound time signature, percussive/non-percussive music, or the kick/snare paradigm. Almost all papers are based on work that makes some underlying assumption about the type of audio that will be subjected to analysis.

Finally, some researchers are concerned with real audio data while others work with intrinsically discrete data such as MIDI note events or sequences constructed from *a priori* knowledge of music theory and the piece under analysis.

C.1 The Reviews

One of the cornerstones of rhythm-tracking research, as judged by the number of times it is referenced in other papers, is Scheirer's 1998 article on tempo and beat detection [1]. Based on the categorization above, this is a fully implemented, frequency-domain, audio-data, engineering paper with a very small set of prior assumptions.

With a basic structure of bandpass filters and comb filters, Scheirer demonstrates non-transcriptive analysis of rhythm using acoustic data of arbitrary harmonic complexity. He proposes that timbre and rhythmic complexity are not correlated. Based on that assumption, he reduces audio to modulated noise, which under certain circumstances carries the same rhythmic significance to human ears. The justification includes a study of past work, all transcriptive and many of them based on segmenting data into onsets. Scheirer argues that human music perception bypasses transcription and onset recognition. This goes against the reviewer's personal experience, as well as the reported experience of music professionals the reviewer is familiar with. Nevertheless, Scheirer's method is attractive because it models yet another mental process percussionists (if not all musicians) use. In the presence of harmonic complexity, onset detection through DSP is error-prone, while human perception is able to relegate harmonic content to the role of background noise when focused on rhythm. This is one of the reasons why percussion instruments are tuned to their best resonant pitches, and not to specific pitches in the key of the music being performed, regardless of the harmonic relationships that ensue. It also goes to show the relevance of studying the findings of the biology and psychophysics of music.

In Scheirer's method, audio data and pseudo-random white noise are run through a filter bank. The envelope of the audio from each filter output is extracted and is used to modulate the corresponding narrowband noise output. (Due to an apparent nonlinear combination of sub-bands in human auditory perception, the noise has to be filtered into sub-bands as well as the audio.)

At this point, the analysis is branched into two parts because, as Scheirer points out, while human pitch recognition is not phase sensitive, rhythm is very much phase dependent. Thus, the rhythmic pulse is defined as having a frequency and a phase, where frequency represents tempo and phase represents downbeat location.

The derivative of the amplitude envelopes is calculated through a first-order difference, and then half-wave rectified. The first step smoothes the audio and the second step reduces sensitivity to false-positives in the algorithm's ability to find note attacks.

This signal is indeed what the trained human brain can automatically do by either listening to the audio, or looking at a waveform or music-score representation.

A separate filter bank of feed-forward comb filters follow each bandpass/envelope/difference stage, where the filter with the delay closest to the input period responds with the highest-energy output. Its frequency is selected as the tempo. The described method of phase determination relies on the particular implementation of the comb filters.

An interesting comparison with time-domain methods points out that while autocorrelation can indicate the existence of rhythmic strata at fractional tempi ($1/3$, $1/2$, etc.), comb filtering can find rhythmic strata at rationally related tempi ($3/4$, $5/8$, etc.). This is a promising result for the study of clave, which necessarily involves fractional, uneven time intervals.

A drawback that became apparent early in Scheirer's paper is that genres with strong beats and simple rhythms were pre-selected. Unfortunately, another significant drawback is that the only audio samples with which the system had no success featured "syncopated instrumental lines and complex drum patterns" and "a mixed, or clave, beat."

On the other side of the spectrum is a demonstrative, time-domain, classical-specific, engineering paper using constructed (non-audio) data by Brown [2] on the determination of meter from note onsets and durations. Not purporting to model human perception, Brown constructs inputs based only on classical European music as test samples. This is because her basic assumption is that "[note] events are correlated from measure to measure." This is an assumption that requires musicological scrutiny.

Using note duration as the weighting factor in the inter-onset interval (IOI) sequence, she claims better results than could be had with equally weighted note events. Interestingly, this could prove to be an advantage in analyzing Latin American

music because while dynamic emphasis tends to be on the first of two closely-spaced notes in European music, it is more commonly found on the second of two such notes in West African-rooted music. A duration-weighted IOI sequence places the emphasis on the second of two closely spaced notes.

Brown demonstrates success with a selection of classical pieces by pointing to peaks in autocorrelation plots. Not only does this involve a human observer to interpret results, but the pieces involve a single melodic line each and are not live performances, or even actual audio. Furthermore, Brown's original method failed with one Mozart piece. Reportedly, "narrowed autocorrelation" succeeds in pinpointing the meter for such "complex" pieces, and a reference is provided for this technique, which the reviewer has yet to look up. Nonetheless, Brown's paper is one of the cornerstones of the autocorrelation trend in rhythm tracking has been motivational in the reviewer's inquiry to date.

Another conference paper that addresses the same technique and statistically compares its many variations reaches a different conclusion regarding weighted and unweighted IOI sequences is Toiviainen and Eerola's [3] comparison of seven types of IOI sequences. The study uses discriminant analysis to rank the predictive success of the seven methods on each of 12,368 European folk melodies, "sampled" at each 16th note.

They found that onset location, not duration, was most important for detecting meter, although they also found that melodic (frequency-domain) content is necessary for detection in the case of isochronous (evenly-timed) melodies.

Focusing on beat hierarchy instead of just the presence of beats (or onsets, attacks, pulses), they demonstrate meter estimation (as with Brown), as well as the absolute location of strong beats (crucial for clave detection). However, this is simply a classification between duple and triple meters, without considering bar length, which again is a pre-requisite for clave detection.

Using the MATLAB¹²³ MIDI Toolbox and calculating autocorrelation for lags of 1 through 16 16th notes, equally weighted IOI sequences performed best. Considering the statistical significance of this paper over that of Brown, and that clave involves very little isochronicity, the idea that duration does not play a constructive role in meter detection through IOI sequences and the 91.5% classification success are considerable.

¹²³ Matrix Laboratory by MathWorks

At the 6th International Conference on Digital Audio Effects in September of 2003, Uhle and Herre presented a paper [4] on estimating certain key features of musical rhythm from pure audio. Using some musicologically interesting definitions, the authors observe that rhythmic structure is not limited to the time signature on a sheet of music, and that notes are grouped into phrases, themes, motifs and other hierarchical structures that can be called “strata.” They also define an original term, Micro Time, to be the ratio of beat and tatum periods. The tatum is the lowest common denominator of rhythmic timing, a uniform pulse series of frequency equal to or the smallest necessary integer multiple of the highest-frequency component of the audio signal.

Tempo, tatum and microtime are used to estimate time signature (where they really mean ‘meter’). The explanation of musicological rhythm categories (additive, divisive, meso-periodic and hybrid) indicates, at the very least, the attempt to apply tempo and meter detection to all possible types of rhythm. For reference, commonly, Western/Northern European rhythm is divisive; Eastern/Southern European and Middle Eastern rhythm is additive; Indian rhythm is divisive or hybrid; and Latin and West African rhythm is meso-periodic (a clave characteristic).

The audio is divided into time segments, each of which is decomposed into frequency bands. Onset detection is achieved via envelope extraction for each band of each segment, which is high-pass filtered and rectified. The tatum is calculated from the inter-onset intervals (IOIs).

In parallel to onset and tatum detection, autocorrelation is used to find periodicity in amplitude. Those autocorrelation peaks that are close to tatum periods are identified as beat location candidates. Integer multiples of those, in turn, are identified as bar length and downbeat location indicators.

The implementation of this is done via the downsampling of audio from 44.1 kHz to 22.05 kHz and reduction from stereo to mono. IIR filter banks ranging from 62 Hz to 8 kHz are used to break the signal into frequency bands. 6th-order elliptic filters with 0.5-dB pass-band ripple and 60-dB stop-band attenuation are used.

The envelopes generated are divided by 43 (why?) and low-pass filtered with a cutoff at 10 Hz. The plot of the resulting signal suggests this is as good a reproduction of the human rhythm perception process as Scheirer’s.

Heuristic criteria are used at the onset detection stage, and a very insightful step of subtracting a linear regression line from the autocorrelation function is used to compensate for the biased autocorrelation estimate in the periodicity estimation stage.

Hypothesis testing is then used for bar length selection (for meter detection) on the trend-corrected autocorrelation function.

Yet another promising model of human rhythm perception is found in Rodet and Jaillet's work on fast attack transients [5]. Based on the common observation that the bulk of audio energy resides in the low-frequency range, the authors focus on abrupt changes in high-frequency content. Instead of wavelets, the authors work with the short-time Fourier transform (STFT) because of lower computational load, even though they state that wavelets have better high-frequency resolution. The STFT is simply a series of time-windowed FFTs.

The working definition of fast attack transient is given as "marked energy peaks ... in several frequency bands." However, this is still a very qualitative definition, as the terms 'marked' and 'several' are not clarified. Also, a mathematical selection was made that does not necessarily correlate with psychoacoustical studies in the choice of exact time of attack when waveforms do not have the perceived infinite slope. The potential mismatch between mathematical and psychoacoustical attack locations is considered insignificant. The so-called aggregation process combines attacks found in different frequency bands. The indicator function generated from the application of a threshold to magnitude and slope information *visually* appears to be a good indicator of percussive attack. However, the sample set of recordings is not as impressive as in Toivianen and Eerola's study, containing 75 recordings, only 17 of which contain sustained sounds. The remaining recordings reportedly contain pure attacks. This is either a potential weakness or an indication that the algorithm was not fully tested at the time of the paper's release for the ICMC conference in 2000.

An important insight from this paper is that a small number of false positives is more important to rhythm detection applications than a higher number of true positives.

The same idea is central to Hainsworth and Macleod's paper on onset detection [6]. Aiming to detect harmonic change without addressing harmonic content, the authors focus on note onsets marked by amplitude changes without corresponding power changes. Using a spectrogram to detect energy increases while ignoring energy decreases (note releases do not contribute nearly as much to a sense of rhythm, except in R&B), a number of distance measures for energy increase from bin to bin are compared.

The best-performing of such measures is then smoothed via convolution with a window and for the time extent between each pair of subsequent crossings of the mean

value, a peak was chosen as the point of sharp energy increase. Some additional elimination is performed because the algorithm performed over-zealously in picking peaks. Though it performed without any false positives, the final algorithm was only tested on only one (albeit challenging) choral sample.

In the section on “Traditional Change Detection,” the authors explain how audio data falls outside of the common constraints for parametric models with additive, uncorrelated noise. Parametric models do not apply because given the varied nature of all music from all cultures and eras, a near-infinite number of models would be needed and model selection would make the task pointless in practice.

An even greater problem in applying statistical assumptions of additive, uncorrelated noise is in the definition of ‘signal’ and ‘noise’ in rhythm tracking. The signal portion of the audio is often inharmonic; i.e., the notes that make up the signal portion from the point of view of rhythm tracking tend to have noise-like spectra. What’s worse, what passes for noise in rhythm tracking tends to be well-behaved notes with harmonic spectra. Thus the problem of looking for a signal in additive noise, as found in communications applications, is turned upside down to finding noise in the signal.

It is the recognition of this dilemma that supports Smith’s treatment of rhythm as the signal, rather than audio as the signal [7]. The frequency of the rhythm signal is $f = \frac{1}{\lambda}$, with λ related to the inter-onset interval and f related to meter.

Since audio information is in the amplitude of the signal (air pressure in sound waves, depth in vinyl LP grooves, 16-bit binary numbers on CDs), it can be treated as amplitude modulation of a carrier in the audible range. Frequency analysis of the signal is used to separate acoustic audio from the low-frequency rhythm content. This is similar to working with MIDI note event data only, and as with Scheirer, lack of timbre (harmonic ratio) information is assumed irrelevant for rhythm detection.

The author selected wavelet analysis for improved time-varying frequency representation than in the STFT. The onset time of each beat is represented as a weighted impulse function. Taking the frequency extent of rhythm to be 100 Hz allows for the exclusion of such complications as flaming¹²⁴.

¹²⁴ Flaming is the percussive practice of separating two simultaneous notes by a barely-perceptible time interval.

The wavelet decomposition focuses in on the high-frequency content in the time domain, while taking longer time segments in the low-frequency ranges. The Gabor wavelet is chosen because it preserves phase in analysis, even though it is not an orthogonal basis function and the signal cannot be reconstructed. The magnitude and phase are plotted with a gray scale for the third dimension. Oscillating shades of gray in the phase plot indicate the presence of a corresponding rhythm frequency in terms of the inverse of the plotted number of inter-onset intervals.

At this point, there is a strange return to the Fourier paradigm. It is unclear why the author makes a statement about Fourier components being in phase where local maxima of energy occur.

The author cautions that monophonic rhythm was used in this study, and that polyphonic rhythm may require parallel wavelet analysis because listeners commonly use timbre and pitch to identify rhythm in the case of polyphony.

Of interest is that the author uses the expression “dynamic and temporal accents,” which is a good alternative to the term “IOI.”

Before making a final statement on the performance of the algorithm, the reviewer wishes to read more of Leigh’s work, some of which is in greater detail than this early paper.

In a related paper with Kovesi [8], Smith tackles expressive timing (feel, swing, rubato, etc.) and rhythmic strata using multi-resolution analysis with Morlet wavelets, decomposing rhythm into short-term low-frequency components that reveal transient details of rhythmic structure.

The authors explain that isochronous rhythm (one with uniform pulse) has a single frequency equal to the inverse of the IOI. Dynamic accenting gives rise to multiple frequencies, hence multiple strata of rhythm, which in turn define meter. They also state that syncopated rhythm (varying IOI, as in clave) may be conceived of as a linear combination of short-term frequency components in a time series.

The term ‘rectification’ is used to describe recovering rhythm from acoustic data. Reference is made to self-organizing neural oscillator. A cursory search of the literature brings up many references to neural oscillators coinciding with references to self-organizing maps (SOM), but no reference to a *self-organizing neural oscillator*. A search of articles on neural oscillators reveals that they are networks of OTAs, somehow arranged or designed to act in an artificial neuron-like fashion. There are also some unclear statements about causality, and that rhythm is conceptual, not perceptual. Also,

the use of British musical terminology (quaver and crochet for quarter note and eighth note) makes the paper a hard read. The purpose of the technique described in this paper is said to be “pre-perceptual processing” which is also unclear to the reviewer at this point.

Dixon and Goebel’s conference paper from the 2007th International Conference on Music Perception and Cognition [9] defines ‘expressive timing’ as “micro-deviations from mechanical timing” as indicated by a musical score. They also suggest that the inter-beat interval is a measure of instantaneous tempo. Their psychoacoustical study concludes that human listeners are not fooled by temporary variations in expressive timing. It is getting machines to not be fooled by temporary changes in tempo that will prove to be the real challenge.

Gouyon, Pachet and Delerue’s 2000 conference paper [10] sets out to classify percussive sounds via zero-crossing rates and a 19-criteria discriminant analysis. Their use of the kick/snare paradigm (applicable solely to Pop, Rock, New Country, Hip Hop and Dance Music) makes their method of little interest to rhythm tracking in other genres. Their sampling frequency of 11025 Hz raises the question of aliasing. Their choice to consider cymbals as noise, when they tend to indicate downbeats very strongly in pop and rock music, appears unwise. Finally, their main conclusion, that decay is more discriminant than attack, is at best applicable to the differentiation of kick and snare sounds, and not likely to be useful for general rhythm-tracking purposes.

Another, apparently unpublished article by Gouyon and Batlle¹²⁵ [11] examines tempo, beat, meter and *rubato* as the four elements of rhythm. The technique involves decomposition into sub-bands and the “transformation” into the “rhythm domain” using a weighted sum of periodograms, which is a very appealing idea to the reviewer. The main restriction in this paper’s approach is that genre has to be known in advance, which is not a drawback for clave detection. Applied only to high-attack instruments, the technique is not fully implemented and requires a human to interpret plots.

Before rounding out this initial review, the reviewer would like to mention another of the cornerstone articles in rhythm tracking. Laroche of Creative (which owns Ensoniq, E-mu, and SoundBlaster) published an article in 2003 [12] on beat and tempo tracking of audio, seeking to pinpoint exact beat locations. Laroche asserts that commercially available software for the extraction of MIDI clock from audio relies on the kick/snare paradigm, and is thus limited only to popular music, possibly only in regular (4/4) time.

¹²⁵ The name is really spelled “Batlle”; not a typo.

Laroche's technique targets real audio of time-varying tempo, but works offline (not in real time). The three steps are the creation of an "energy flux" signal from the audio signal, calculating a host of tempo and beat candidates and making a selection.

Laroche argues that frequency-domain analysis is better than time-domain analysis of signal energy because loud, long notes (whether from an electric bass in rock music or the pedal note in an organ fugue) can mask changes that signify beats.

After performing STFTs with non-overlapping, 10-ms windows, each frame's FFT is compressed to bring out high-frequency percussive attacks and pitch changes, and suppress high-amplitude long notes in the low frequencies.

In order to find energy increases, first-order differences from frame to frame are calculated and all bins are summed and half-wave rectified. The resulting signal has sharp peaks at transients and onsets.

At this point, a predetermined set of tempos and beat locations are compared with the rectified energy differences. In order to do this with the discrete FFT outputs, tempo and downbeat locations are made discrete as well. Each tempo value R_i falls into one of N_R discrete tempo values. For each possible tempo, each downbeat location, t_j^i , falls into one of N_D discrete time values. So there are multiple downbeat location candidates at every fraction of the tempo candidate. This is the pre-constructed, discrete expectation signal.

A set of cross-correlations is calculated between the actual and expected for each tempo/downbeat combination. A local search algorithm is used to make the final selection for each frame. A tempo track is also generated, which maps out the tempo variation as derived from the instantaneous tempo changes based on first-order beat location differences.

The algorithm does not find downbeat locations with much success. Since clave detection requires the establishment of an absolute reference or starting point, this is a very significant drawback.

The stated inability of the algorithm to choose beats 1 and 3 instead of beats 2 and 4 for samples of rock music, however, is strangely welcome for clave detection, which may commonly require music that also has a strong beat on 2 and 4. This shortcoming can be taken advantage of by redefining the downbeat from the musicological sense to the folkloric sense.

Laroche's paper on tempo, swing and beat location estimation from the 2001 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics [13] is

based on the assertion that commercial software for extracting a MIDI clock from audio signals “fail miserably” when working with any type of input other than techno music, and that more generally applicable solutions are needed. The use of rhythm-tracking methods in audio editing software is emphasized and a specific example of the use of rhythm tracking in *cut-and-paste* operations is given.

Laroche makes the initial assumption that input files will have constant tempo. This assumption is one of the typical trade-offs that the reviewer needs to consider in his research; the answer is not as obvious as it is with some other trade-offs.

The method described has two parts: transient analysis (extraction) and maximum-likelihood detection. The definition of transient used by the author is “times at which the energy of the signal in some frequency band increases sharply.” The absolute value of the short-time Fourier transform (STFT) is used as the energy of the signal in the frequency domain. The energy content in each frequency band is summed, and the first-order differences on the energy function from one time frame to the next are calculated. The peak extraction method that follows is said to work for non-obvious high-frequency content as well as wide-band percussion hits.

‘Swing’ is defined by Laroche as a slight delay of the second and fourth sixteenth notes. The reviewer disagrees with this definition, as the second sixteenth note is almost never played in standard jazz swing or blues shuffle. How would it be possible to know where it falls in practice? Furthermore, there are many different swings, such as jazz swing, samba swing, reggae swing, *hardingfele* swing, and *taiko* or *dentou geinou* swing. At least some of those do not follow this description. Moreover, in defining the probability distribution for note events around the swing/beat grid, the location of the first beat, the percentage amount of swing, and the track duration are inputs. However, these are also the very same quantities that are being estimated, and should be considered unknown to the system at this point.

In addition, there is no theoretical or practical support given for the selection of a pseudo-Gaussian shape for the probability distribution.

As yet another questionable choice, only one techno track is used to generate a histogram of beat strengths to determine a priority between the four sixteenth-note beats. Especially since the author stated that the applicability to only techno music was a shortcoming that this paper attempted to undo, the statistically unsound selection of a single techno track for determining beat strength priorities is more disturbing than simply making an informed decision based on personal knowledge of music would be. However, the author does state that the assumption is false for other genres, and since

the *Techno* genre is relatively homogenous, this assumption is noted, along with the others.

In the implementation, the log-likelihood of a transient falling on any one of the sixteenth-note beats is maximized through an exhaustive search. The author admits that deriving the *pdf* from actual observations is a “chicken/egg problem” because it would require knowing beat locations, tempo and swing amount. Two of these were already assumed in the selection of the *pdf*, as stated above. Considering the influence that this paper has had on future papers in beat tracking and beat induction, the order and reasons behind these assumptions may need further scrutiny.

Further *ad hoc* choices are made regarding the variance of the distribution and how it is adjusted by the tempo. The author argues that a tempo grid of $1/8$ BPM is necessary to make sure hypothesized beats line up with actual beats. The argument for this is unclear to the reviewer.

In order to speed up the performance of the tool, a small portion of the audio is deemed sufficient for tempo estimation under the assumption that tempo is constant. This is an acceptable assumption for some genres of music when recorded in a digital sequencer or studio environment, but will fail to hold true for recordings of live music or expressive studio recordings of classical music and some other genres. In order to account for expressive timing, the author explains that the first estimate of the tempo can be updated with a subsequent search on the entire signal with “a smaller search space centered on the [initial] estimate” [13].

Further revisions are made on the *pdf* based on a lack of dependence on the first beat location. The results indicate consistently wrong beat location in Latin American music, and no result or false results in the absence of percussive or sharp peaks. For varying tempi, the author suggests using overlapping time frames and updating current estimates based on previous ones, “possibly in a Bayesian framework” [13].

The lack of success with Latin American music in this questionable but central paper on rhythm detection supports the need for the reviewer’s pursuit of a Brazilian-specific approach to rhythm tracking.

Some years prior, at the 1995 IJCAI Workshop on Computational Auditory Scene Analysis, two prominent names in rhythm tracking from Waseda University presented another influential paper on real-time beat tracking [14]. The paper promises a beat-tracking method for “realistic acoustical environments” and makes the interesting and important statement that beat tracking is essential as an initial step in

computer understanding of music, specifically in the case of Western (northern) music, and particularly for pop and rock music. The tool described in the paper is said to recognize the time positions of beats (quarter notes) in real time, but requires “context-dependent” knowledge to resolve ambiguous beat locations. The basic assumptions are 4/4 time signature (acceptable for most instances of clave-direction analysis), 65-185 notes per minute, almost-constant tempo (but not absolutely constant—potentially acceptable for clave analysis), and that drums maintain the beat (not necessarily true for clave-based music).

Such assumptions are a common feature of most work on rhythm detection of any kind; assumptions based on knowledge and understanding of musical context are central to attempts in computer recognition of rhythmic elements. The determination of context is one aspect of so-called automatic detection that may prove to be the most challenging in the development of rhythm-tracking methods. In other words, human input may be required to make the initial assumptions and perhaps also to select an appropriate method of analysis at all times.

The method of [14] uses pre-recorded drum patterns in memory for matching against perceived patterns in the signal under analysis. Multiple “agents” using different strategies to test a variety of hypotheses are employed, and implemented on a parallel computer for real-time performance. The analysis is divided into a “low level” and a “high level.” At the low level, onset times are determined. At the high level, these onset times are interpreted into beats. The system is said to have performed successfully for 40 out of 42 pop songs analyzed.

The issues to be addressed are listed as: 1) ambiguity in beat location; future data may be needed to determine beat locations; 2) pure frequency analysis is not enough; musical knowledge is used to make decisions in the selection of beats; 3) not all interpretations will be correct; ongoing adjustment of parameters may be needed, as no set of initial parameters is appropriate to all inputs.

To handle the first issue, multiple agents are set to examine different frequency bands, more than one per band, using different hypotheses. (The reviewer would like to suggest different time intervals as well as frequency bands.) Also, each agent evaluates its own reliability based on the performance of its beat location predictions in upcoming stages.

To handle the second issue, eight common drum patterns are placed in memory to differentiate between strong and weak beats, as opposed to simply seeking out quarter-note subdivisions. However, it is important to note that this locates the

study firmly in the “kick/snare paradigm” described in part 1 of this literature review, which is not applicable to the study of Latin American or West African music.

The rate of change in signal power, the power present in neighboring time-frequency values, and the prediction of future beats are used to evaluate the reliability of onset estimates to address the third issue.

Furthermore, communications between the agents is implemented to adjust or eliminate low-performing agents, based on a measure of reliability.

Beat information is defined as a combination of beat time and beat type, where type-Indicates strong and weak beats, based on the kick/snare paradigm of strong bass drums hits on the downbeats and weak but noticeable snare hits on the backbeats. For analysis of clave-based music, it should be noted that this is at times the exact opposite of the beat structure found in Latin American and West African music , in which strong, low-frequency beats fall can on the backbeat (such as in samba) or on no strong beat at all, as identified in the European idiom (such as in *Sabar Woloff*).

In the frequency analysis portion, the algorithm detects onsets in a number of frequency bands and determines the onset times of bass drum and snare hits. The power spectrum is calculated via FFT with overlapping Hanning windows. Half of the CD sampling rate, 16 bits and a 1024-sample window size are chosen. At quarter-window overlap, this gives a frequency resolution of 21.53 Hz, and a time resolution of 11.61 ms. Roughly a hundredth of a second would translate to 6000 notes per minute, which would correspond to an effective rate of faster than 128th notes at 120 BPM. This is certainly a sufficiently fine grid, as few humans are capable of really playing 128th notes at any reasonable tempo. (This reviewer cannot imagine anyone other than virtuoso percussionist Evelyn Glennie executing such a thing to any degree of accuracy.)

Fast-rising power components are extracted by simple comparison to the two immediately neighboring times and the two such frequency bands. Similarly, noise components are extracted for snare detection. While the perception of snare sounds as noise is interesting, snares are not likely to be a vital part of clave recognition, so the reviewer does not focus on this aspect of the paper. For the bass drum notes, onset times are compared with known bass drum characteristics; i.e., the peak with the lowest frequency in the histogram is the detected bass drum. The reviewer believes some sort of threshold may be needed to make that decision reliable.

The author notes that in real-time, onset times are well in the past as their analysis is completed, which is a central lesson for musicians to learn: mistakes are part

of the past, and dwelling on them simply increases the chance of future mistakes. Similarly, the system is onto new onsets by the time a given onset has been evaluated. Therefore, many decisions have to be made in parallel if the system is not to fall irrevocably behind. This is the reason for the parallel implementation. Also, there is a complex interaction between the agents that may be beyond the reviewer's scope.

An interesting and promising insight from [14] is that the IBI (inter-beat interval) is selected according to the interval with the maximum peak in the IOI (inter-onset interval) histogram, which is weighted by the reliability of each agent.

Though the paper includes such valuable insights and presents a creative solution to the decision making problems of rhythm tracking, its dependence on the kick/snare paradigm and the reliance on parallel implementation make its current use to the reviewer somewhat limited. However, a non-real-time version to be implemented in a non-parallel fashion, with adjustments for genre-specific needs in clave detection could prove useful.

In a related paper presented in the year 2000 and sponsored by the Austrian Federal Ministry of Education, Science and Culture [15], Dixon and Cambouropoulos tackle beat tracking based on context-based information, but this time allowing for large tempo variations. However, they deal with MIDI data only, skipping the analysis steps necessary to extract and choose onsets and interpret them as beats. Also, the method is applied only to classical piano sonatas. Once again, the authors emphasize the value of such research for intelligent audio editing and next-generation search engines.

The authors define beat induction as finding the tempo (or tempi), and beat tracking as finding beat locations (in time). They refer to music-theoretical studies indicating that, in addition to the simple identification of onsets, the dynamic hierarchy of notes is critical to finding the beat. The dynamic hierarchy ("accent structure") underlies the "relative salience of musical events ... determined by ... duration, dynamics, pitch, harmony and cadences" [15]. This is an important contribution to rhythm tracking, and music perception in general, because it reflects and models human perception and understanding of music. Not all notes (note onsets) are equally salient.

The authors make the reasonable assumption that any interval of less than 0.025 seconds is not an inter-onset interval. Such short intervals could be unintentional or expressively intentional but not meant to be beat-related. They also consider that inter-onset intervals could exist between note events that have other note events

between them. This is another important notion that the reviewer has not encountered elsewhere, but which in retrospect, fits the notion of *tatum* (to be explained below) and how it relates to the beat.

A ceiling of 2.5 s was selected as the limit for inter-onset intervals with interfering onsets. Using a set of rules, all the IOIs are assigned to one of many clusters, differentiated by time lengths. The number of IOIs is found for each cluster and constitutes that cluster's score. The highest-scoring IOI clusters are candidates for the beat interval and its subdivisions and multiples. These are awarded extra points in their scores (the exact scheme is not given), and a final rank order is determined. The authors imply that about a 20-s sample is enough to determine tempo for pop music (though they don't explicitly say 20 seconds).

In a parallel with one of the papers reviewed earlier, the authors express the rhythm as a signal with frequency and phase (the rhythm domain). In terms of this rhythm-as-signal approach, frequency (tempo) has been discussed so far, but not phase (beat location).

An agent is created for each hypothesis, similar to Goto and Muraoka's method [14], for each of the first few onsets and each of the first few locations (phases) until the combinations are exhausted (though the definition of "few" is lacking). Each agent has a current state and a history of selected frequencies and phases. Each agent is evaluated for regularity and salience.

Each agent predicts locations for the entire audio input duration. The main program passes each new onset found to each agent as a time value. Under given tolerances, the agents compare the onset to their predicted beat locations. There are two tolerances, inner and outer, with the inner being small and symmetric in time, and the outer varying with the inter-beat interval as it is calculated in real-time. If an onset meets the inner tolerance, the agent is updated as having predicted that beat location. If only the outer tolerance is met, a new agent is created under the current agent. Redundant agents are removed.

When an accepted onset is more than one interval away from the last event of that agent, based on the agent's particular interval size, the intervening spots are artificially filled. This is said to allow for tempo variations and syncopation—though the authors don't use that term—and is remarkably similar to the process the author of this literature review goes through when listening to rhythm.

The salience value used in [15] is primarily harmonic. While this could be used in clave detection, the harmonic contribution to the salience of notes in clave-based

music needs further clarification and study. On the other hand, a time-based salience structure can be identified more readily using the handful of characteristic rhythmic frameworks in Afro-Brazilian music.

The authors note that system performance increases from around 75% to around 90% when salience is introduced to the decision-making process, applied to 222 audio selections from 13 Mozart piano sonatas.

Again, though applied to a completely different type of audio signal, [15] provides many rewarding insights, such as the assignment of agents to different beat-interval candidates, the method of filling intervening values, allowing for tempo variation by creating sub-agents with the outer tolerance, and most importantly, the use of intervals across more than one onset.

On a different note, a 1965 IEEE paper [16] describes a nonparametric detector for signals of known form in the presence of non-Gaussian noise. Since the portion of an audio signal that is defined as noise (from the point of view of rhythm-as-signal) is rarely likely to be Gaussian, this paper is of great interest to the reviewer.

The paper discusses signals of exactly known form, and signals of partially known form with unknown phase. No knowledge of noise distribution is assumed.

There are n independent observations of an m -dimensional vector, in *iid* noise from a continuous CDF ($F(y)$), otherwise unknown. $\mathbf{X} = \xi\mathbf{C} + \mathbf{Y}$, where \mathbf{C} is the known signal and \mathbf{Y} is noise.

The i -th coordinate of \mathbf{X} has the CDF: $F(x - \xi C_i)$. The detector decides whether $\xi = 0$, or not.

Each dimension of the n observations can be combined so that

$\mathbf{S} = [x_1^{(1)}, \dots, x_n^{(1)}, \dots, x_1^{(m)}, \dots, x_n^{(m)}]$, where $x_j^{(i)}$ is the i -th coordinate of the j -th observation.

Instead of a matched filter that correlates the received vector with the known signal, under the assumption of additive Gaussian noise, the nonparametric detector uses hypothesis testing, where the null hypothesis is that $x^{(i)} = y^{(i)}$, or that the received signal is pure noise, with arbitrary distribution type, mean and variance.

Restricting ξ gain values to low signal-to-noise ratios, first, the $N = nm$ received samples are ordered from low to high amplitude.

$E[Z_{(k)}]$ is the expected value of the k -th largest of N samples from a zero-mean/unit-variance Gaussian (!) population. These values are stored at the detector. The received samples and the stored expected values are partitioned into “ m classes of n samples each” —as many as there are signal dimensions, or coordinates.

In each class, the expected values are summed and the sums are scaled by C_i . The resulting value is compared to a threshold. If greater, the null hypothesis is rejected and detection is positive.

The authors explain that this scheme is indeed very similar in reasoning to the matched filter, but simply makes fewer assumptions and does not use absolute amplitude information, only relative amplitudes based on ordering the input samples. The multiplication by C_i is particularly similar to the matched filter, as it reinforces the sample values received when the signal is present and not when the signal is absent.

Performances of the method were compared to the matched filter in both Gaussian and other noise using Pitman’s Asymptotic Relative Efficiency (ARE) as the criterion. The ARE of one procedure relative to another is “the limiting ratio of sample sizes required by the two detectors to yield the same error probabilities for a decreasing sequence of gains (ξ)” [16].

The theorem put forth in this paper is that the ARE of the nonparametric detector as compared to the correlation detector is unity under Gaussian noise, and strictly greater with other noise distributions. In other words, the nonparametric detector of a signal of known form performs better under non-Gaussian noise than a correlation detector. If verified, this could be a useful result for musical audio.

In another rigorous article that is much cited in the rhythm-tracking field, a previously reviewed author, Laroche examines spectral analysis for highly damped sinusoids with close frequencies [17]. His paper focuses on three characteristics of musical audio signals and the specific frequency analysis needs that result from them: high spectral resolution, ease of use, and robustness. Laroche notes that musical signals, even when generated by single instruments playing solo, contain two or three very close damped sinusoidal components, which beat against each other. In some cases, it was the failure of electronic instrument manufacturers to take into account these neighboring frequencies around the pitch-defining frequency that contributed to the noticeable artificiality of electronic instruments. Hence, it is now understood that these beating frequencies are critical to the achievement of natural sound, and therefore to the perception and analysis of real musical signals. This requires a high degree of spectral resolution. Robustness and ease of use come into play when an analysis

method is not so sensitive to an initial choice of analysis parameters that it fails to converge for less-than-excellent parameter choices.

The paper elaborates on Hua and Sarkar's spectral method published in the IEEE ASSP journal in 1990 [18]. Interestingly, a clear distinction is not made between this method (called the matrix pencil method) and the very similar ESPRIT method co-authored by Kailath and Roy in 1989 [19].

The signal is assumed to be made of exponentially decaying sinusoids: $x[n] \approx x_n = \sum_{i=1}^L A_i e^{-\alpha_i n} \cos(2\pi n f_i + \phi_i)$, with frequencies in cycles/sample, $f_i \in [-0.5, 0.5]$, and related to the "real frequency" F_i by $F_i = F_s \times f_i$ and a similar relationship between the sampling frequency and the damping factors.

After making an assumption on the number of sinusoids in the signal, L , a very specific set of matrix manipulations are prescribed, based on two matrices created from x_n 's. These calculations are said to give the frequencies and damping factors of the assumed component damped sinusoids. Theoretical development was left to the references cited.

In the second iterative step, a new set of matrices and vectors are formed from the damping values calculated above, and the amplitude and phase values are calculated. It is noted that the calculations require finding the Moore-Penrose inverse, which involves the SVD (singular value decomposition), which in turn requires Eigenvalue Decomposition. It is also possible to have an implementation where the Moore-Penrose inverse is calculated directly from the Eigenvalue Decomposition. However, especially in this latter implementation, though really in both versions, the computational cost increases rapidly with the so-called "pencil parameter", p .

The choice of the parameters p , N , and L affect the accuracy of the spectral decomposition. While the assumption is that the signal is purely a sum of damped sinusoids, this is not necessarily precisely true. As a check, Fourier analysis can be performed on the signal, and peaks above a threshold can be counted to determine L . Overestimating L is not as bad as underestimating it. This gives a working range for the parameter p , such that $2L \leq p \leq N - 2L$. Higher values of p trade off computation for spectral resolution. Therefore, p is essentially the parameter of resolution of close frequencies. Choosing p requires extensive knowledge of the signal (another chicken/egg problem) in terms of the number of pairs of beating frequencies and the SNR. However, if we know the number of pairs of beating frequencies, is the current analysis not moot? This is a question that requires further inquiry.

It is stated that the parameter p cannot be greater than 500 because the matrices in the recipe suffer from numerical instability. This implies that this method is impractical on signals of more than 2000 samples. Therefore, not only is a low sampling rate recommended for long-decaying sounds, it is also crucial to select a portion of the signal after the note onset (decreasing energy).

This is potentially a major shortcoming, or at least a computationally expensive requirement, because all other methods encountered by this reviewer feature an emphasis on the rising-energy (pre-onset) portion of signal sections. The stated advantage of this method over the FFT for spectral analysis is the use of very short windows (which are in fact, a requirement). Once the necessary assumptions are made, the method gives the signal parameters without assuming specific peak locations. It is also said to be more robust and easier to use than, and with comparable resolution to the Prony and MUSIC (Multiple Signal Classification) methods of spectral analysis.

All in all, Laroche's method appears more suitable to identifying percussive instruments present in an audio signal (content analysis) than to analyzing the temporal evolution of amplitude peaks (rhythm detection). Therefore, in the very advanced stages of automatic detection, a method such as this could come in handy for eliminating the front-end human factor in classifying audio based on instrumental content.

On a different note, the reviewer's familiarity with the now-crucial term *tatum* began when reading Seppänen's paper on tatum-grid analysis [20]. The tatum is related to the smallest time interval found in a piece of music. However, it is not exactly, and simply, the smallest time interval. A *flam* can feature a time interval between two notes that is so small that the two notes are considered one (in terms of musical analysis). Thus, there is a lower limit on the time intervals that can be considered for tatum. Furthermore, the actual tatum interval can be perceived or induced, rather than heard, if the interaction of various beats in the rhythm hierarchy is such that a smaller interval is implied without actually being performed. While rare, this is a valid possibility. Possibly for these reasons, Seppänen gives a definition of tatum in terms of its relation to the beat hierarchy: "the pulse intervals on all other metrical levels ... are integral multiples of the tatum" [20].

In addition to the polyphonic/polyrhythmic interpretation of tatum, another valuable insight in this paper is the implementation of hysteresis in the selection of onsets: Amplitude must drop below a certain level that is lower than the level it has to rise above to be considered an onset, before rising above the latter.

An onset detector is employed that detects changes in the amplitude envelopes of frequency sub-bands and flags as onsets points of rapid increase in amplitude. An exponentially decaying past-data window is used to allow for tempo variation. This is an interesting idea, but seemingly rather arbitrary. For any selection of the rate of exponential decay (and what about other forms of decay?), there may be a realistic tempo change option that would not be fully attenuated or over-attenuated. Also, the reviewer believes that if tempo variation exists, it would be necessary to (somehow) detect tempo variation and remove it by leaving it outside a rectangular window, as continuous changes in tempo would wreak havoc on the accurate calculation of tatum. In other words, tatum is only meaningful at a given tempo. The exponentially decaying window is at best a first approximation of how tempo variation should be treated in the interest of tatum measurement.

In any case, eight non-overlapping logarithmically distributed frequency bands (eight sub-bands) from 45 Hz to 19 kHz are created using nonlinear phase filters for computational efficiency. The audio samples tested include the genres of Jazz, Rock, Techno, Classical, Big Band and Pop.

The RMS amplitude of each sub-band is estimated and the sampled signals are decimated to 100 Hz. Each is then convolved with a half-raised cosine window in order to model human hearing. A scaled first-order difference is then compared to a threshold to determine whether an onset is present or not. Once onset detection is completed, all sub-bands' onsets are combined into one set of onsets for the full audio input.

As in [15], not only are first onsets taken, but also overlapping onset values up to a ceiling value. This is because tatum is not necessarily just the shortest audible time interval, as explained above, but the shortest *inducible* time interval.

In the absence of tempo fluctuations, the tatum is the greatest common divisor of the inter-onset intervals. In the presence of tempo variations (deviations in the inter-onset intervals from expected values), a remainder error function is defined. Calculating the remainder error function for arbitrary tatum-period and IOI values resulted in an error value that did not make sense to the reviewer.

The equation is:
$$e(p) = \sum_{i=1}^n \left(\frac{o_i}{p} - \left\lfloor \frac{o_i}{p} + \frac{1}{2} \right\rfloor \right)^2.$$

Let's assume a tatum period of 3 and an IOI of 5.5. The floor of $5.5/3 + 1/2$ is 2. The error contribution for this IOI value becomes $1/36$.

In order to ascertain whether there is a pattern that makes sense, let's try a tatum period of 1 and an IOI of 3.7. The floor of $3.7/1 + 1/2$ is 4. The error contribution becomes 0.09.

It is clear that this result is not what was intended by Seppänen. However, one does not know what value to plug into p because this discussion of the remainder error function is the first mention in [20] of a period p . Hence, it is completely unclear what p is a period of. This is a serious problem in a paper published by the IEEE.

In a further attempt to allow for tempo variations, IOIs are converted to a series of histograms of IOI-per-second, evolving over time and updated at every onset. The remainder error is calculated on the histograms and the tatum is chosen from the local minima of the errors.

The reported success of the system is mixed, with the bulk of the failures in the most expressive pop, jazz and classical pieces. While the author claims that tatum is a good starting point for mathematically calculating meter, the incomplete explanation of the implementation and the spotty success rate do not make this a very attractive method.

The final paper discussed in this review is on the use of subspace analysis for drum transcription [21]. When separating sound sources combined into a single channel (or two stereo channels, which can be combined into a single channel, and which provide little source separation when left separate), the amount of information necessary for blind source separation depends on the input signal. In this paper, sub-band processing is used to overcome the indeterminate factor of the arbitrary input signal. This method of source separation, first proposed in 2000, reduces redundancy in signals by interpreting sound sources as sub-dimensions of the time-frequency representation of the signals. In order to do this, assumptions are made about the sound sources that caused the observed signal.

An STFT is performed on the signal. Coefficient magnitudes of the STFT are the spectrogram for n frequency channels and m time slices. In an $n \times m$ matrix Y , each column vector is the frequency spectrum for a given time slice j , and each row is a frequency sub-band evolving in time.

It is assumed that the spectrogram is made up of l independent unknown spectrograms, Y_j and that each Y_j can be represented by the cross product of a time-invariant frequency basis function, f_j , and a similarly invariant amplitude envelope, t_j . $Y_j = f_j t_j^T$.

At this point, the reviewer has difficulty agreeing with certain claims made by the authors. The first of these is that stationarity implies no pitch change. However, the concept of statistical stationarity has nothing to do with the behavior of the point statistics of a population drawn from a distribution; it does not simply mean “non-changing values”. If the authors are claiming that pitches are drawn randomly from a statistical distribution, this is not clear in the paper. Furthermore, only in a certain limited set of genres may this statement be true.

Secondly, the authors say that drum sounds are stationary in pitch. Again, if stationarity is taken to mean not-changing, the reviewer strongly disagrees that drum sounds are stationary in pitch. It is common knowledge, at least among drummers, that professional drum set toms, for example, are often tuned with one lug out of tune with the rest, so that the drum exhibits a falling pitch as it decays. Similarly, Indian *tabla* and West African talking drums feature highly changing pitch, as do certain other drums from Peru, Brazil, Korea and Japan (e.g., the *tsudzumi*) to name a few.

If all these assumptions were to be taken as valid, the main issue would be the estimation of the basis functions for the sound sources. The spectrogram is decomposed with Principle Component Analysis, via the Singular Value Decomposition; $Y = USV^T$ where U is an $n \times n$ orthogonal matrix, S is an $n \times m$ diagonal matrix of singular values, and V is an $n \times m$ orthogonal matrix. The columns of U are the principle components based on f , and the columns of V are the principle components based on t . p , the number of sources, is much less than either m or n .

The first few principle components are kept as the independent basis functions. However, Principle Component Analysis does not do that. Independent Component Analysis is necessary for independent non-Gaussian sources to be separated. A linear combination of such sources is assumed: for an observed signal mixture \mathbf{x} , a vector \mathbf{s} containing independent non-Gaussian sources, and a mixing matrix \mathbf{A} , $\mathbf{x} = \mathbf{A}\mathbf{s}$.

It is worthy of note that the authors’ use of the Central Limit Theorem does not take into account the potential for heavy tails in non-Gaussian distributions.

If \mathbf{A} mixes \mathbf{s} sources into \mathbf{x} , then they are assuming that the \mathbf{x} signals have PDFs that are “more Gaussian,” and the \mathbf{s} sources, less so. This is not a terribly rigorous argument. By this reasoning, an inverse to the matrix \mathbf{A} is to be found that gives a “very non-Gaussian” set of signals, and the independent sources (basis functions) have been recovered. The corresponding time and frequency basis functions are obtained by multiplying the original combined signal spectrum with the pseudo-

inverses of the independent basis functions. At this point, the lack of rigor with which this theory is presented has made it less than a promising method of source separation.

The use of source separation in audio would be in pre-processing to simplify onset detection, or the selective removal of instruments whose parts traditionally do not contribute to a sense of clave direction. As indicated in the conclusions, many improvements are needed in this model to remove the limitations and loosen the constraints presented.

In “Pattern Discovery Techniques for Musical Audio,” Dannenberg and Hu [22] start out with a number of important observations about the nature of music and music perception. One such observation is that recorded music is the most common and least informative way of storing and representing music. Other representations, such as formal notation, informal notation, or MIDI data contain explicit information about the melodic, harmonic, rhythmic, timbral and structural aspects of the piece of music. Audio, on the other hand, contains no organizational information whatsoever. Therefore, it is when applied to pure audio that music recognition is most pertinent. And in fact, there is more to music recognition than rhythm tracking. The authors remind the reader that such techniques as “key detection, chord identification, [...] transcription, melody and bass line detection, source separation, speech recognition [for lyrics], and instrument identification¹²⁶” are continually being developed to describe music at higher levels than pure audio, and that the ultimate unification of all these measures and techniques is sought in order to develop expert systems that can characterize music at a meta-level.

While high-level characterization of music is the ultimate goal of all these types of music recognition, some degree of high-level analysis is also a prerequisite to each music recognition scheme. According to Dannenberg and Hu, musical structure is indicated by repetition at some structural level higher than that of digital audio samples, or even individual notes. While the thrust of the paper is melodic and harmonic structure, the same concept applies closely to clave, as clave direction arises from a mid-level abstraction of the time-energy relationship of notes in a piece of music.

The authors use John Coltrane’s ‘Naima’ as the signal under analysis, and focus on describing the common jazz harmonic structure known as ‘AABA,’ wherein the piece of music has two harmonically-different sections (sections arranged around a different tonal center) and the so-called B section provides a form of relief from the

126 The reviewer would like to add genre detection to the list, which is a highly active area of research in MIR.

main, A, section. It is duly noted that the techniques described in this paper are not those of rhythm tracking, but the concepts may still be applicable.

In reviewing related work, the authors mention other pattern-search efforts. One cited work is a researcher's attempt to describe the characteristics of classical composers [23]. Another cited work examines the relationship between data compression and music description [24] in the sense that both focus on the repetition and transformation of patterns.

[22] proceeds to present a number of approaches to defining and comparing patterns, starting with a brute-force approach of comparing all possible samples monophonically. Not surprisingly, this method is found to be computationally impractical. In a second attempt, Bartsch and Wakefield's concept of *chroma* [25], which locates like segments, is explored. The chroma is an expression of tone color in vector form. There are 12 elements to the vector, each representing the total energy of one of the 12 tones of the European octave. If the audio signal is segmented in time, a chroma vector can be generated for each segment and a distance measure between vectors can be defined after normalizing the vectors. The segments are of equal duration. For some reason not clearly explained, the matching of such segments is declared ambiguous. Dynamic programming is suggested for finding the shortest path between chroma vectors as a way to define similar segments. This is still too expensive. A heuristic is proposed that similar segments should follow diagonal paths. Diagonal paths are said to be similar in tempo. The implication is that the energy in a given frequency (and its octaves) is a function of tempo. Higher tempi imply lower energy, as each peak has less time to complete its contour. Why this is found in diagonal motion through the matrix is not explained. It is also necessary to note that the authors do not explain how exactly the matrix is formed from the set of chroma vectors (rows or columns?) until after the reason for the choice of heuristics is given.

There are three matrices; one for distance, one for path, and one for length, called **D**, **P**, and **L**, respectively. A set of computations is defined, which require more knowledge of dynamic programming than the reviewer has gleaned by reading online tutorials. Mysteriously, the authors then switch to yet another algorithm for similarity: polyphonic transcription without source separation. This means that while simultaneous but different notes are identified and transcribed, they are placed in the same track/instrument. A polyphonic neural-network piano transcription program, SONIC [26], and a chordal analysis program, Harman [27], are used. The authors then go on to explain an algorithm combining polyphonic transcription and dynamic programming. This is a search algorithm based on a measure of similarity. The goal is to find a set of similar time segments. All in all, it is very difficult to tell the objective of

[22]. Without stating which algorithm is used, the authors move on to clustering the results of the previous algorithms, where they assert that the monophonic and *chroma* techniques succeed in defining correct high-level structures for jazz, classical and popular recordings, though they only give one example in each category. While they have clear success with the Beethoven minuet, their attempt to derive the ‘AABA’ structure in ‘Naima’ seems contrived, at best. (See figures in [22].) The reviewer can only interpret the resulting plot as “AABCADEFFFEBAACC.” This is far from ‘AABA’ even after removing the FFF section, which is revealed to be an improvised solo.

Another interesting work is that of Alghoniemy and Tewfik on “periodicity detection” [28]. This work is interesting in two ways: It makes some important psychoacoustical observations, and at the same time, shows a significant lack of musical knowledge by the authors. First of all, they chose to use a song by ABBA (which they identified by name), and a song by The Beatles, which they referred to only as a “children’s song,” not knowing that it is in fact a psychedelic story by The Beatles.

Secondly, they state that “beats are found in the low frequency region of music signals,” which is categorically incorrect for both *Bata* and *Ketu* drumming in Cuba, Brazil and West Africa, and as the authors admit, not necessarily true for European classical music either. (In certain religious and traditional musics from Nigeria, Benin, Ghana, Cuba, Uruguay, and Brazil, bells or high-pitched drums “keep the beat” (provide the grounding *ostinati*) while the lowest pitched drum takes on the creative, improvisational, expressive role.)

Thirdly, the description of what may or may not be *accelerando* and *ritardando* with expressions such as “when beats gets faster, we notice a drop in the time-difference” leads to ambiguity because one cannot be sure whether the authors refer to changes in tempo or changes in the rate of change of tempo.

Last but not least, in the section on “different cultures,” there is an implication that the envelope of the low-frequency beat-defining instruments found in Indian and Egyptian music is related to the use of odd and compound meters in those cultures. While one may be able to claim that musical traditions can be said to evolve in response to environmental factors, which in turn affect instrument construction, which determines timbre and envelope, making such an assertion would require ethnomusicological and psychoacoustical backing.

Having addressed these issues, let’s consider some of the valuable insights and contributions from this paper. In the introduction, the authors talk about the psychological comfort found in grouping events such as beats. Charles R. Hoffer [29] is

quoted as saying that humans tend to hear a grouping of “tick-tock” sounds coming from a wall clock, when in fact the sound is “tick-tick-tick-...” The reviewer himself has frequently conducted the informal psychoacoustical experiment of listening to such regular sounds as can be found in the home, and mentally grouping them into 2’s, 3’s, 4’s, and even 7’s. Amazingly, one out of every n sounds, where n is the number of sounds per grouping, is perceived as being different (“tock” rather than “tick”), regardless of what value of n is chosen. A sympathetic accent is synchronized to the perceived strong beat and the act of hearing takes into account both the observation and the expectation. This is sometimes called “subjective rhythmization” [30].

Subjective rhythmization necessarily raises the question of how valid any choice of grouping is when unchanging patterns (ostinati) are employed, especially in music where no cultural framework exists (as it does in European music) for differentiating time signatures such as 2/4, 4/4, 4/8 or 8/8. (See below.)

For example, in so-called “4/4” rock and pop music, there are two strong low-frequency beats per measure. Does this mean that the time signature should really be 2/4, with only one strong low-frequency beat? Is the choice of four quarter notes (in 4/4 time) merely for convenience? After all, the average phrase last longer than a measure of 4/4, and the common accent is more frequent than 4/4. Perhaps the true time signature for pop music is 8/2.

Here we see that different musical traditions interpret the ideas of meter and time signature (the statement of meter) differently. According to London,

In counting according to one meter and not another, a musician gives a series of tones a particular rhythmic shape and nuance; their sense of the meter leaves a kind of residue in performance, such that the “same” series of notes played under different counting frameworks will have distinctive differences in its expressive timing and dynamics. [31, p.4]

This notion of meter is certainly valid for the music that European standard music notation was developed for (and under). But does it apply to the musics of other cultures? There certainly are very specific preferences for “expressive timing and dynamics” in Afro-Brazilian music (those aspects called *suingue*, *cadência*, and *balança*), as well as in the rest of the Diaspora. However, it is clear from the different ways these musics can and have been notated that meter—in African Diasporan contexts—is more of a choice of phrase lengths, which sometimes gets traded off against visual convenience: The rhythmic phrase in samba is almost always four beats long, but “jazz cats” are used to reading in groupings of two. Hence, samba gets notated in 2/4 in the

US (and either 4/4 or 2/4 in Brazil, depending on the author's interest in conforming to US jazz convention.) Some observations of the educational literature in the US and Brazil for Brazilian, Cuban, West African and Uruguayan music are in Appendix J.

Musicologist Danielsen's synthesis of the work of Ghanaian musicologist Kwabena Nketia and ethnomusicologist Arom on meter strongly supports our claim that meter signals only phrase length (Ekwueme's *okele* [32], which is Arom's *isoperiod* and Danielsen's "*basic unit*" [33, p. 43]) in African and Afro-Latin music, and not the accent structure associated with meter in the European conception.: "Arom emphasizes that this periodicity occurs through the repetition of an identical or similar unit of musical material and not through giving some beats more weight than others. [...] He claims that accentuation in isoperiodic music does not form a regular accentual matrix: every pulsation or beat within the period has the same status. [...] no matrix of strong and weak beats" [33, pp. 43–44].

Hence we see that *samba batucada* is most appropriately notated in 4/4 time, which reflects the vast majority of phrase lengths, and not 2/4 time, as seen in many US publications whose emphasis is on the recurrence of strong and weak onsets.

Moreover, returning to London's explanation quoted above, this is likely to be the underlying reason for the distaste Brazilian, Cuban and other Diasporan master musicians (who are generally not conservatory-educated) typically have for the rhythmic expression of highly trained, highly accomplished European or North American musicians who may have seen the written in European notation more often than they have heard it played, and who have internalized the European convention of what is indicated by meter.

In addition, it is likely that the tick-tock psychoacoustical effect and the expressive nuance arising from meter [31, p. 4] are one and the same mental process.

Returning to the review of MIR literature, the beat-tracking approach proposed by Alghoniemy and Tewfik [28] starts with narrow low-pass filtering. The amplitude signal is visually observed for beats. This reviewer believes that when the authors refer to "beats getting faster," they mean that finer subdivisions of the beat are being employed, not that the tempo is increasing. The occurrence of finer subdivisions, probably in the form of drum fills, is taken as an indicator of repetition segment starts.

At this point, it is clear that this paper does not describe an algorithm. It does not discuss a full implementation. It is of a descriptive nature, where the conclusions result from human interpretation of plots, not machine execution of a decision-making

process. This is notable because it means that the IEEE publishes papers on rhythm perception even when they are only descriptive (essentially incomplete).

The second section is on periodicity analysis. Large- and small-scale patterns are sought in the amplitude of the audio. The beat peak threshold was found to be about 70% of the maximum amplitude, though it is not explained what this is based on.

Input signals are treated differently based on whether or not they have sections where the apparent beat (indicated by amplitude peaks) stops. A threshold is taken halfway between the signal minimum and maximum, and all signal points are expressed as -1 or $+1$. Binary tree parsing is used to identify patterns following each run of either value. While the reviewer cannot currently give a true review of the methods used, he has started reading about Lempel-Ziv and trellis parsing in data compression as a result of reviewing this paper, and may be able to give a more in-depth review in the near future. This particular pattern finding idea using tree parsing of the audio samples seems to be a promising candidate for decision making based on the cross-correlation sequences between the known and observed clave patterns¹²⁷.

The paper ends with the statement that real-time implementation of the pattern-finding algorithm is possible, meaning it had not been implemented at the time of publishing. Nonetheless, the introduction to data compression is appreciated.

In another attempt at characterizing rhythmic patterns, Douglas Eck of the Dalle Molle Institute for Artificial Intelligence in Switzerland proposes a modification and combination of previously published positive- and negative-evidence models of patterns matching [34].

The idea is very simple. The rhythm is matched against uniform beat sequences of whole, half, quarter, eighth and sixteenth notes. The role of melody and dynamics is not included in the simple pattern-finding rules, so the application is somewhat limited. Also, the method described is likely to be fooled into choosing an offbeat as a downbeat because it doesn't see the forest for the trees.

The simple set of rules described in this paper are probably applicable to rhythms in European music, but a completely different set of rules would need to be developed empirically, based on musical knowledge, for clave-based forms. For example, rule three says that the "first and last of three or more adjacent beats are

¹²⁷ This idea was pursued by the researcher prior to the current dissertation, and discarded as inconclusive and ineffective.

accented.” This is clearly not true in some Brazilian and West African music, where the middle of three adjacent beats could easily carry the accent.

In any case, under the given assumption, a pattern of beats is determined and “quantized” (in the musical sense). A set of clocks is generated from the accented beats selected according to the simple set of rules. All clocks have periods equal to or less than half the period of the pattern, and are even-numbered divisors; i.e., half, a quarter, perhaps a sixth, definitely an eighth, etc. The number of times each clock coincides with three types of beats, accented, unaccented and rest, are counted.

The negative-evidence models rate the clocks based on the numbers of unaccented beats and rests that coincide with clocks. It has unfortunately been shown that this model does not perform well under tempo variations.

The positive evidence model rates clocks based on the number of accented beats that coincide with each clock, which will necessarily give the highest rating to the fastest clock. The normalized positive-evidence model proposed by the authors makes up for the deficiency of the original positive-evidence model, by dividing each score with the frequency of the clock being matched to. (Actually, it was stated as multiplying by the period.)

This makes sense. There is some interesting treatment of tempo variation that does not make that much sense at the moment. The primary goal of the paper seems to be to compare the model’s behavior with that of two classes of humans, musicians and non-musicians. The time-varying normalized positive-evidence model is said to perform satisfactorily especially at higher tempi when compared with musicians, though a hybrid of the positive- and negative-evidence models also had good matches with musicians’ choices in the low tempi.

The last paper in this review is a rating and ranking of several dozen statistical measures in terms of their success rate in predicting genre differences based on tempo [35]. The concept that genres can be determined through tempo appears utterly ridiculous, until one notes that the authors only target music for ballroom dancing. It is important to note, however, that even under this limitation, it is not the musical genre that is being identified; it is the dance genre. The claim that there is no difference between samba and rumba other than a change of tempo is beyond ridiculous and could only imply that the genres under consideration are not musical genres. True samba and rumba in music are not limited to the tempo ranges specified in this paper; they can overlap in their tempo ranges. The real difference between samba and rumba

are in their fundamental patterns, their types of swing, their instrumentation and their characteristic clave patterns.

Even more interestingly, this paper claims that the difference a waltz and a rumba is also purely tempo. Anyone with a modicum of music knowledge knows that a Waltz is in triple time and a Rumba (*Guaguanco*, for example) is not. The claim that tempo is “of prime relevance in classifying musical pieces” is simply absurd. The only worthwhile statement is in the conclusion, and is to the effect that “time signature, swing and syncopation” should also be considered for genre identification. The reviewer would argue that instrumentation would also be a major factor.

C.2 Conclusion

There is much more to the audio DSP and MIR literature than discussed here, and since it is an active area of research, new techniques, new applications, and new ideas are continually being developed. The articles reviewed above include some of the cornerstone works in MIR, and some emerging ideas at the time of proposal of the present research. Since then, onset detection, tempo tracking and genre recognition have continued to attract interest, and developments have been made using techniques ranging from non-negative-matrix factorization to Kalman and particle filters.

In conclusion, the main obstacle to intelligent processing, classification, and retrieval of music seems to be *context*, which manifests itself in two challenges:

1. Mathematical and engineering methods typically require the identification of conditions and assumptions for which a given technique can be successfully applied. In music, vast differences in tempo and instrumentation combined with the timbral richness of almost any sound source has so far kept techniques that are fruitfully applied to one subset of music from being similarly deployed in other musical contexts. In contrast, in speech, meaning is paramount, rather than timbre, pitch or rhythm. (Those are important even to meaning, but such changes in speech do not affect what is being communicated nearly as radically as in music.) Hence, as important as statistical and adaptive techniques of signal processing are to music research, intelligent discernment of context is the central challenge. A human listener with minimal cultural awareness can locate a beat or track a tempo with much greater success than any machine solution to date.
2. Both as a result of the contextual challenge and as seen in the preceding review of the literature, domain expertise is a necessary (though not

sufficient) requirement for music research. Ideally, all MIR research would employ a music specialist (or several), a technical specialist in signal processing, a technical specialist in machine intelligence, and a specialist in Statistics. The first of these is sometimes missing, and the fourth is almost always missing. It is the conviction of the present author that one way or another, all four areas of knowledge and inquiry must be present to make technological research in music relevant.

References for Appendix C

- [1] Scheirer, E. D., "Tempo and beat analysis of acoustic musical signals," *J. Acoust. Soc. Am.*, Vol. 103, No. 1, 1998, pp. 588–601.
- [2] Brown, J. C., "Determination of the meter of musical scores by autocorrelation," *J. Acoust. Soc. Am.*, Vol. 94, No. 4, 1999, pp. 1953–1957.
- [3] Toivianen, P., and Eerola, T., "The Role of Accent Periodicities in Meter Induction," *Proc. 8th Int. Conference on Music Perception and Cognition*, 2004, pp. 422–425.
- [4] Uhle, C., and Herre, J., "Estimation of Tempo, Micro Time and Time Signature from Percussive Music," *Proc. 6th Int. Conference on Digital Audio Effects (DAFX-03)*, 2003, pp. 1–6.
- [5] Rodet, X., and Jaillet, F., "Detection and Modeling of Fast Attack Transients," *Proceedings of the International Computer Music Conference (ICMC'01)*, 2001, pp. 30–33.
- [6] Hainsworth, S., and Macleod, M., "Onset Detection in Musical Audio Signals," *Proceedings of the International Computer Music Conference*, 2003, pp. 163–166.
- [7] Cemgil, A. T., Kappen, B., Desain, P., and Honing, H., "On Tempo Tracking: Tempogram Representation and Kalman Filtering," *Journal of New Music Research*, Vol. 29, No. 4, 2001, pp. 259–273.
- [8] Smith, L. M., and Kovesi, P., "A Continuous Time-Frequency Approach to Representing Rhythmic Strata," *Proceedings of the Fourth International Conference on Music Perception and Cognition*, 1996, pp. 197–202.
- [9] Dixon, S., and Goebel, W., "Pinpointing the Beat: Tapping to Expressive Performances," *Proceedings of the 7th International Conference on Music Perception and Cognition (ICMPC7)*, 2002.
- [10] Gouyon, F., Pachet, F., and Delerue, O., "On the Use of Zero-Crossing Rate for an Application of Classification of Percussive Sounds," *Proc. COST G-6 Conference on Digital Audio Effects (DAFX-00)*, 2000, pp. 1–6.
- [11] Guaus, E., and Batlle, E., "Visualization of Metre and Other Rhythm Features," *Proc. IEEE Symposium on Signal Processing and Information Technology*, 2003.

- [12] Laroche, J., "Efficient Tempo and Beat Tracking in Audio Recordings," *J. Acoust. Soc. Am.*, Vol. 51, No. 4, 2003, pp. 231-232.
- [13] Laroche, J., "Estimating tempo, swing and beat locations in audio recordings," *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 2001, pp. 135–138.
- [14] Goto, M., and Muraoka, Y., "Music understanding at the beat level—real-time beat tracking for audio signals," *IJCAI-95 Workshop on Computational Auditory Scene Analysis*, 1995, pp. 68–75.
- [15] Dixon, S., and Cambouropoulos, E., "Beat tracking with musical knowledge," *Proc of the 14th European Conf on Artificial Intelligence*, 2000, pp. 626–630.
- [16] Daly, R. F., and Rushforth, C. K., "Nonparametric detection of a signal of known form in additive noise," *IEEE Trans. Inf. Theory IT-11*, 1965, pp. 128–134.
- [17] Laroche, J., "The use of the matrix pencil method for the spectrum analysis of musical signals," *J. Acoust. Soc. Am.* Vol. 94, No. 4, 1993, pp. 1958-1965.
- [18] Hua, Y., and Sarkar, T. K., "Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise," *IEEE Transactions on Acoustics, Speech and Signal Processing*, Vol. 38, No. 5, 1990, pp. 814–824.
- [19] Roy, R., and Kailath, T., "ESPRIT – estimation of signal parameters via rotational invariance techniques," *IEEE Transactions on Acoustics, Speech and Signal Processing*, Vol. 37, No. 7, 1989, pp. 984–995.
- [20] Seppänen, J., "Tatum grid analysis of musical signals," *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 2001, pp. 131–134.
- [21] Cemgil, A. T., Kappen, B., Desain, P., and Honing, H., "On Tempo Tracking: Tempogram Representation and Kalman Filtering," *Journal of New Music Research*, Vol. 29, No. 4, 2001, pp. 259–273.
- [22] Dannenberg, R. B., and Hu, N., "Pattern Discovery Techniques for Musical Audio." *ISMIR 2002 Conference Proceedings: Third International Conference on Music Information Retrieval*, M. Fingerhut, ed., Paris: IRCAM, 2002, pp 63–70.
- [23] Cope, D., *Experiments in Musical Intelligence*, Madison, Wisconsin: A-R Editions, Inc., 1996

- [24] Simon, H.A. and Sumner, R.K. "Pattern in Music," in Kleinmuntz, B. ed. *Forma Representation of Human Judgment*, Wiley., New York, 1968
- [25] Bartsch, M. and Wakefield, G.H., "To Catch a Chorus: Using Chroma-Based Representations for Audio Thumbnailing," *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 2001.
- [26] Marolt, M., "SONIC: Transcription of Polyphonic Piano Music With Neural Networks," *Workshop on Current Research Directions in Computer Music, (Barcelona, 2001)*, 2001, pp. 217-224.
- [27] Pardo, B., and Birmingham, W. P., "Algorithms for Chordal Analysis," *Computer Music Journal*, Vol. 26, No. 2, 2002, pp. 27-49.
- [28] Alghoniemy, M., and Tewfik, A.H., "Rhythm and periodicity detection in polyphonic music," *Proc. IEEE 3rd Workshop on Multimedia Signal Processing*, , 1999, pp. 185-190.
- [29] Hoffer, C. R., *The Understanding of Music*, California: Wadsworth, 1989.
- [30] Bilmes, J., "A model for musical rhythm," *ICMC Proceedings*, pp. 207-210. Computer Music Association, 1992.
- [31] London, J., *Hearing In Time: Psychological Aspects of Musical Meter*, New York, NY: Oxford University Press, 2004.
- [32] Ekwueme, L. E. N., "Concepts of African Musical Theory," *Journal of Black Studies*, Vol. 5, No. 1, 1974, p. 45.
- [33] Danielsen, A., *Presence and Pleasure: The Funk Grooves of James Brown and Parliament*, Middletown, CT: Wesleyan University Press, 2006.
- [34] Eck, D., "A positive-evidence model for classifying rhythmical patterns," *IDSLA tech report*, Dalle Molle Inst. for Artificial Intelligence, 2000, pp. 1-11.
- [35] Gouyon, F., Dixon, S., Pampalk, E., and Widmer, G., "Evaluating rhythmic descriptors for musical genre classification," *AES 25th Int. Conf.*, 2004, pp. 1-9.

APPENDIX D: Statistics, Social Responsibility, and the Enhanced Scientific Method

The core of science is not mathematical modeling; it is intellectual honesty. (Sam Harris, *Beyond Belief*, 2006)

This appendix brings together all the theoretical and philosophical reasons behind the choices of statistical and experimental design that were made during the defining phases of the research for this dissertation, and endeavors to explain why correct and appropriate statistical analysis is critical to research in *any* field. Included is a focused coverage of relevant aspects of the field of Statistics as well as a broad justification of how and why it was used in the present work. It also serves to answer a question asked of another doctoral candidate at another institution: “[W]hat distinguishes you, in the way you’re doing this research, from a common hired laboratory technician?” [1, p. 131]

The difference is in the broad scientific, philosophical and cultural—or simply, *systems*—approach of this work. Such an approach differentiates scholarship (even in applied fields) from the routine execution of experiments. Among the differences are:

- the design of relevant experiments through
 - the continual design and rejection of earlier sets of experiments and
 - the continuous inquiry into scientific and statistical methods, and
- attention to the technological and cultural context of the problem and its areas of application.

D.1 Why Statistics?

Statistics the field (differentiated from the plural of a statistic) is not a branch of Mathematics. Mathematics is exact and internally consistent. Mathematical results follow with certainty from a small set of fundamental axioms, and even when theorems remain unproven, dispute regarding their nature or significance is not of the deep philosophical nature as with disputes between schools of Statistics. For instance, the debates between frequentist, axiomatic, and Bayesianist statisticians are as deeply rooted in the nature of meaning and physical reality as is the debate between the proponents and opponents of M-theory in Physics. Mathematics, on the other hand, possesses solved problems and unsolved problems, not philosophical debates about whether they ought to be tackled at all. In this sense, Statistics is more like a Science or a form of applied mathematics. This fundamental difference between Mathematics and Statistics is

also reflected in the separation of departments of Mathematics and Statistics into disparate academic units at most leading institutions of higher education¹²⁸.

Statistics is, instead, the *science of Science*. It provides both the theory and methods for carrying out scientific investigation. Specifically, it protects the experimenter against the effects and ubiquity of noise, the *cum hoc, ergo propter hoc* fallacy, and the presence of confounding effects such as nonlinearity and interaction [2, pp. 7–9].

On the other hand, Statistics is not yet a mature science. There are many outstanding questions, ranging from the degree of applicability of techniques developed for the social sciences to the physical sciences, to the significant challenges trained scientists face in applying, interpreting and representing statistical analyses, and on to the rift between fundamental interpretations of the meaning of probability and its consequences in statistical data analysis. No doubt similar statements can be made about competing theories in physics, chemistry or genetics, but those fields are relatively mature in that certain successful, reliable practical results follow from widely accepted theories enjoying substantial evidence. In Statistics, however, *fundamental* principles still seem to be under debate.

Nonetheless, the importance of applying Statistics to scientific and technological research, and the importance of doing so with a critical eye toward the relevance of various techniques are recognized throughout the sciences. The importance of Statistics is self-evident; the importance of recognizing the aforementioned problems is validated by a growing body of research into the misuse of statistical techniques.

D.2 The Role of Statistics in the Scientific Method: Statistics as Meta-Science

The material in this section is based primarily on literature research in Statistics conducted late 2009 through early 2011, and presented at the Portland State University Systems Science Seminar Series¹²⁹ on February 25th, 2011, titled *Some Problems & Solutions in the Experimental Science of Technology: The Proper Use and Reporting of Statistics in Computational Intelligence, with an experimental design from Computational Ethnomusicology* [72].

This expanded discussion is intended to explain and justify research decisions regarding data selection and preparation, and the core experimental design. The topics addressed include the scientific method (past and present, inductive or hypothetico-deductive), critical thinking, the misuse of statistical techniques, hypothesis testing and

¹²⁸ Examples include UC Berkeley, University of Michigan, Stanford, Harvard, Carnegie Melon, Texas A&M, Penn State, University of Washington, Purdue, UCLA and the University of Wisconsin.

¹²⁹ Presentation slides available: <http://www.pdx.edu/sysc/seminar#Winter2011>, recording available: <https://sas.illuminate.com/site/external/jwsdetect/playback.jnlp?psid=2011-02-25.1246.M.6D598EC9FA3A19B4F7B15EDAC00EB6.vcr&sid=2009262>

associated problems, techniques of data selection for experimental design, and application of these issues to the present data and experiments.

As discussed previously in this document, sampling a rather small subset of all possible occurrences of a natural, social, or technological process is often the best we can do in terms of gathering data for analysis. Problems of induction aside, such data, in turn, help us make predictions or draw conclusions about the mechanisms underlying complex processes.

Not only is our ability to observe severely limited, but so is the process of making measurements.

Through these two sources, natural systemic errors (devoid of any blame or responsibility) enter into our observations. The job of scientists, and in fact, all living beings, is to make the best possible inferences with the data at hand.

Statistics is a science developed around notions of probability for the purpose of improving our ability to make decisions, predictions, inferences, instruments, and develop an understanding of the universe and ourselves. Statistics bears its own limitations in the very language used to express its findings: Certainty is unlikely. Uncertainty rules. Within these limits, we aspire to make better judgments and inferences, and also attempt to avoid drowning in the degrees of uncertainty necessarily attached to each uncertain observation or conclusion.

In short, we cannot know all there is to know, so we sample instead. In sampling, we build models of the state of nature as it really is (unknowable). These models come in two types: structures and hypotheses. Collected data (observations) come together to suggest a structure, and we put forth hypotheses to explain the underlying mechanisms, such as gravitation, evolution, consciousness, or the economy.

In order to judge the relative merits of our hypotheses, we test them through further observations. (We will avoid deteriorating the discussion into such issues as the realism of the inductive versus hypothetico-deductive approaches, confirmation or falsification until later.) Such testing is necessary because the errors introduced by the finite nature of our data collection (not to mention biases unwittingly introduced into the data) make it so that *there is no way to know with absolute certainty* whether the mechanisms we posit or the predictions we make reflect the underlying truth of the system we are investigating.

Nonetheless, statistical rigor is not hopeless or useless, because we can establish degrees of certainty and confidence (that are themselves subject to uncertainty, as Type-2 Fuzzy Sets are compared to regular Fuzzy Sets). Since we have to conduct our daily lives by ignoring extremely unlikely outcomes and paying attention to highly plausible ones, there is no practical reason not to seek higher levels of confidence in our scientific endeavors (save for Hume's critique of inductively justifying induction).

The practical question, then, is: What is hypothesis testing, and why is it important?

Whether we are asking questions about a process improvement from one design or method to another, or seeking correlations among observables, we are always pitting the possibility that something desirable or exciting may be the case against the possibility that nothing interesting is going on. Because of the necessary emphasis on skepticism in Science (see Sections D.3 and D.4 below), the latter, status quo, is considered the norm, also called the *null hypothesis*. For example, in a comparison of algorithms, the null hypothesis would state that there is no difference between the different algorithms' performances. In a drug study, the null hypothesis would be that the drug is ineffective. In seeking a correlation, the null hypothesis would be that there is no correlation to be found. This is because in Science the burden of proof that a new truth has been uncovered is on the scientist, and sufficient reason must accumulate before a new idea can be accepted.

To make matters worse, remember that in no case do we actually know the underlying truth. This means that if we are to make claims about the efficacy of a drug, a policy, or a neural network (where lives, in each of the three cases, may be at stake), the duty of the scientist is to raise the bar significantly higher than in any other pursuit.

Hence, in addition to betting on the null hypothesis, we only allow a very slim chance of rejecting that the null hypothesis is true, typically a 5%, 1% or smaller likelihood.

For starters, let us assume that this small percentage is the chance that we would be wrong if we were to reject the null hypothesis and claim that our design or prediction is indeed an improvement over the prior state of affairs.

Since we can never know for sure, the best we can do is to accept only small chances of being wrong, which we choose smaller as the stakes (such as lives) get higher.

Furthermore, to insure scientific honesty, we set and declare our target likelihood (small chance), called α level, in advance. This is the highest (worst-case) probability of being wrong about rejecting the null hypothesis that we're willing to risk (because lives, money, or knowledge is at stake). Then we wait and see if our data and calculations indicate a chance of being wrong in rejecting the null hypothesis that is smaller than the target we set. (This is the calculated p value.)

If $p < \alpha$, we declare *STATISTICAL SIGNIFICANCE*.

There are several problems with this, as we shall discuss below. But first, consider that the difficulty comes not so much in calculating that likelihood, but in interpreting what it really means.

For starters, let us take the textbook description: “whether the observed difference between the samples might be attributable to chance or sampling variation rather than to real differences between the populations” [3, p. 80].

This is an important question to ask because all studies are based on partial samples (of all cancer patients, all flight-traffic controllers, all single-hidden-layer neural networks, all archaea, all voters, etc.), and no matter how many techniques we use to eradicate various influences (biases), we neither can know that we did so, nor that such biases do not exist in the true population behavior. So-called sampling error is omnipresent, and that is why hypothesis testing is essential: We cannot make absolute claims (no matter how much the public misinterpret scientists’ refusal to “sound confident”). We can only present our findings along with the chances of them not reflecting the underlying truth of the system under test.

However, even this careful approach has its pitfalls. Assume that a study with a target statistical-significance level of 5% resulted in a p value of 0.02. We would interpret this to mean that if we were to reject the null hypothesis and declare a positive result, there would be a 2% chance that we would be wrong.¹³⁰

That sounds like the probability that the null hypothesis was indeed true would be 2%. This is where the typical scientist ends and the statistician begins, for the statisticians tell us that this is simply not the right interpretation [4, p. 184].

If that is not the meaning of p , then what is? Miller offers the following:

P-values are correctly defined as: the probability of the observed data or data more extreme, given that the null hypothesis is true, and the sampling was done randomly. [4, p. 185]

Since Statistics is based on Probability, the truth of the null hypothesis, according to the above statement, is *given*. Note that Miller’s statement does not say “given the null hypothesis,” but says “given that the null hypothesis is true.” It would appear that we know the null hypothesis to be true. Carver’s interpretation has a similar ring to it: “Statistical significance [simply] means statistical rareness. Results are ‘significant’ from a statistical point of view because they occur very rarely in random sampling under the conditions of the random hypothesis” [6, p. 381].

Let’s play devil’s advocate and look at this from a point of view different than the usual:

¹³⁰ What does that really mean? In reality, we are either wrong, or not. That underlying truth exists, though it’s not reachable. There really is no likelihood of being wrong in rejecting the null hypothesis; we are either right or wrong.

1. If calculations of statistical significance are to have any meaning, their assumptions must be true. If their assumptions are true, then the null hypothesis is true.
2. If the null hypothesis is true, then results as extreme as obtained (this is the standard vocabulary) are significant, which really means rare.
3. Hence, we have found rare results that are not correct (because the null hypothesis is correct).

If this line of reasoning is correct, then the “decline effect” [18] is a no-brainer: The statistical significance levels found in initial studies decline to not-significant levels when studies are repeated (this is an observed phenomenon in the life sciences) because the initial finding was a rare and false finding (which got published due to its statistical significance), and later studies naturally cannot reproduce the effect.

If the above interpretation seems like nitpicking on a technicality, consider Carver’s statement that “[t]he p value, which can only be calculated by assuming the truth of the null hypothesis, is what researchers use to decide whether or not to reject the truth of the null hypothesis” [6, p. 381]. If the null hypothesis is true, then, necessarily, the alternative hypothesis is not, in which case why even bother with the significance (rarity) of what has to be *false* results?

If the null hypothesis is *not* true, as suggested by the p value, then the assumption for calculating the p value is violated, and we cannot trust the calculated likelihood.

It seems hypothesis testing is logically bankrupt.

A possible objection is that hypothesis-testing calculates *how wrong we are to make the opposite claim* when the null hypothesis really true. However, “how wrong we are to make a claim” is a useless statement. “The likelihood that we are wrong” would make more sense, but that raises the possibility that the null hypothesis may not actually be true, even though the hypothesis-testing procedure operates under the condition that it is true. This is a vicious cycle. It seems there is a mistake, perhaps at the *Probability* level, before we get to Statistics.

To make things worse, considering the issue from the point of view of an engineer or a scientist, Miller makes the observation that “[t]he null hypothesis is known to be false before the test is conducted” [4, p. 186]! This is because, in many cases, a two-sided null hypothesis will state (absurdly) that the true mean is equal to some exact, specified value. Under a continuous probability distribution, any exact value has a probability of zero.

Even if we can brush aside this concern by saying that some scientists do not set up good (i.e., interval) null hypotheses, we have not gotten around the problem that many null hypotheses are constructed to satisfy the hypothesis-testing process. Proper

scientific inquiry should have all parameters defined in advance. Tailoring assumptions to a numerical process that will assume those assumptions is another vicious cycle, and appears not to be good science.

Finally, Bayes theorem (which is a fundamental fact of Probability) indicates that the probability of a hypothesis is not equal to the probability of experimental results. (This, again, relates to the base-rate fallacy.) So, even if hypothesis testing were otherwise meaningful, its results should not give us the level of confidence scientists and the media typically express with regards to statistically significant findings.

There appear to be problems with both the theory and the practice of hypothesis testing. Further inquiry in this direction requires a review of the history of how the practice came about (which raises further issues because the intentions of Fisher do not seem to have been fully compatible with those of Neyman-Pearson (cf. [5, p. 156])).

Nonetheless, there also seems to be practical value in hypothesis testing (cf. [2]). After all, we started out with the fact that we have to sample populations, and that this introduces unknown and unavoidable errors. It was reasonable to ask how much of a risk we would take by trusting our sampled data. After all, in voting studies, for instance, the vote will eventually take place, and it will either turn out like the sample result or not. In that case, we soon find out whether the null hypothesis was right or wrong. To say that it has to be true—that the universe is a certain way—because *we carried out an unrelated calculation* is absurd.

If our data showed that candidate A would win, and candidate A does indeed win, having made the *P*-value calculation does not alter reality.

The practical usefulness of hypothesis testing, statistical significance, and related concepts is in urging researchers to move asymptotically closer to the point where the sample is the population, and the outcome is the reality. Without this thrust, scientists would soon behave like regular people; that is, reach conclusions about complex systems based on a sample of one.

While the present author continues to learn about the issues surrounding hypothesis testing, the practical concerns of its intentional and unintentional misuse and misreporting are here considered more poignant than the philosophical underpinnings and their misinterpretation.

(The preceding section does not read like a literature review, but indeed it is. The literature on statistical significance, statistical power, sample size and hypothesis testing is such that the interconnections among the articles and the concepts are more important than picking each article apart one at a time.)

At this point, it is best to move on to the misuses of hypothesis testing and the misinterpretation of statistical-significance *results* (not the idea itself).¹³¹

D.2.1 Misuse of Statistics in Science, Medicine and Technology

[J]ust because someone with a Ph.D. or M.D. performs a clinical trial doesn't mean that the trial possesses any credibility whatsoever. In fact, the vast majority of these efforts are worse than worthless because they produce misleading results. [7, p. 102]

Assuming statistical-significance testing is a valid and meaningful thing to do, there are several ways in which it goes wrong in common practice:

1. Ignoring multiplicity effects
2. Publication bias
3. The rule of the arbitrary cutoff
4. Low power and sample size
5. Conflating statistical and practical significance

It is all too common for scientists, engineers and other professionals to learn some basic statistical techniques, and use these indiscriminately under different circumstances.

Freedman *et al.* have reviewed 717 articles in medicine to find only 33 randomized controlled trials (already a cause for concern), and found that only three of those included a calculation of sample size for statistical power, and the remaining studies reported statistical findings or interpretations without having had the sample size to detect the effects discussed [8]. Ioannidis famously declared that “most published research findings are false” [9] because of effect size versus sample size, and other concerns. Likewise, Siegfried stated that “if you believe what you read in the scientific literature, you shouldn’t believe what you read in the scientific literature [10]. The apparent dilemma is not just humorous. Although using scientific studies to criticize the execution of scientific studies may have Humeian problems of inductive justification (rather like the relativist fallacy discussed below), the fact remains that scientific practice often does not live up to its own claims of rigor and objectivity. Some of these issues are further addressed in the ensuing scientific discussion in response to Ioannidis.

¹³¹ This is not to say that the “case against statistical significance testing” [6] is irrelevant or closed, but it simply falls too far beyond the scope and application of the present work.

Ziliak and McCloskey contend that more than 80% of articles in *leading* journals equate statistical significance with importance [11].

Similarly, Cohen, a leading figure in studies of statistical power, observes that most scientists ignore power analysis and sample-size calculations (especially *a priori*) [5]. Salzberg gives the striking example of a study where 14 classifiers were compared on 11 data sets [12, p. 320]. Each of these was compared to a default classifier using a two-tailed paired *t* test with $p < 0.05$. The problem is that this is not the correct test for this study. With this setup, there is *at least* a 99.96 % chance of incorrectly claiming statistical significance:

There are 154 chances for a result to be statistically significant. Thus, the expected number of significant results is $154 \times 0.05 = 7.7$. To calculate the proper α value for this study, let's first define α^* as the probability that if there is no true difference, we find at least one statistically significant difference.

Then, $1 - \alpha^*$ is the chance of getting the right conclusion per experiment.

This, in turn, raised to the n th power is the *optimistic* chance of making at least one mistake. This is *optimistic* because the *t* test assumes independence, so the calculation above is valid under an assumption of n independent test sets.

The real alpha value is $\alpha = 1 - (1 - \alpha^*)^n \approx 0.0003$. In other words, any p value higher than 0.0003 should not be taken to indicate a statistically significant result. And if n distinct test sets were not used, the true α target is even lower.

The statisticians have provided us with various tools, such as the Bonferroni adjustment to the pairwise comparison, or the Dunnett and Friedman ANOVAs, to deal with cases such as these. The problem is that they are numerous and confusing. The truth of the matter is that every scientist has the responsibility to learn and understand their application.

Publication bias is the unfortunate but natural situation that arises when many scientific studies on the same topic (all assumed statistically sound) are submitted for publication to a small number of prestigious journals during a short period of time. If four out of ten studies discovered statistically significant results at a given target rate (that is common to all studies, for the sake of argument), and six did not, it is likely that some or all of the four will be published, and some or all of the six will be rejected.

Gould explains the phenomenon thus:

Only the most miniscule proportion of scientific studies ever get reported in the press, and these decisions often bear little correlation with the importance of such studies for professionals. Better relationships can be found between the decision to report and the degree to which a conclusion disturbs conventional notions (often misconceptions) about the nature of things. [74, p. 208]

A more telling situation is if more than twenty such studies were carried out, and only one resulted in findings at a statistically significant level. When that lone study gets published, the act of publishing that study is itself not statistically significant at the typical 5% level¹³².

Negative results are not appealing to the human psyche. But they are just as important. Knowing that statistical significance does not correlate fully (if at all) with importance, it is imperative for the healthy functioning of the scientific endeavor for negative results to be published.

The rule of the arbitrary cutoff is when we reject findings just above the arbitrary cutoff (of, say, 5%) or wholeheartedly endorse results just below the same [13, p. 1457; 4, p. 186]. In fact, statements of statistical significance themselves may prove not to be statistically significant [14]. (This is like designing Type-3 Fuzzy Sets, otherwise known as a serpent eating its tail.)

If the target level was 1%, and the p value obtained was 0.00101, the study should not necessarily be discarded as “insignificant” because the target was set arbitrarily. It is an honesty guideline, not a physical law. There is nothing in nature that says all processes, from biological to economic to social to technological should have an exact cutoff of 5%.

There are, however, good arguments in favor of sticking with the typical cutoff [15]: Would an athletic team, a company, or a university let each applicant set their own thresholds for achievement in order to be accepted into the team, company or university? Certainly not. No matter how arbitrary, standards are set universally. Berger and Hsieh have hit upon an excellent analogy in their defense of the universal statistical-significance target. Nevertheless, using expert knowledge and intuition to interpret the resulting p value is key in correctly using statistics: Carry out the statistical procedures correctly, but then set them aside and make decisions based on your domain knowledge.

This goes back to publication bias, wherein good studies that do not show statistically significant results are barred from reaching the rest of the scientific community, possibly leading to duplicated efforts and likely leading to only having the studies that show statistical significance get any exposure. This gives the public and the scientific community a lopsided, even wrong, impression of the underlying reality.

¹³² See <http://xkcd.com/882> for a fictional, humorous demonstration of this.

Another important concern is statistical power. This is the other side of the coin from statistical significance: While α concerns the probability of a type-I error (no true underlying difference, but a difference is observed in the sample), the sample size necessary to guarantee a statistically significant likelihood of discovering or capturing an effect of a given size (expressed in terms of fractions of variance) depends on the β target. Statistical power concerns type-II errors, which in some cases [16, p. 2] are more critical than type I. A type-II error means there is no underlying difference between candidate solutions, but a statistically significant difference was observed in the sample nonetheless. β is the probability of making a type-II error, and $(1-\beta)$ is called statistical power. Since β is typically set to 0.2 or less, this implies an 80% chance of detecting a difference. It is important to note that we have just switched from talking about falsely finding a difference that is not there to *detecting* a difference that is perhaps small enough to require statistical “power” to reveal. This is the same type of troublesome switch that happened in the discussion of statistical significance with the notion of denying a null hypothesis which is assumed to be true *and* known to be false.

Nevertheless, for the pragmatic reason that to use these techniques is better than living even more to chance, we also need to pay attention to statistical power.

Fortunately, statistical power can be calculated given the type of test needed (at the end of the experiment), the significance level α , the effect size of interest (80, 50, or 20 percent of the variance) and one of the two following: sample size or β . Knowing the sample size, we can calculate our statistical power; knowing the desired power, we can calculate the necessary sample size.

Some outstanding problems in this area include an ongoing debate that the standard sample sizes developed by Cohen [5] are not valid outside of social science. In fact, for the present study, the required sample size of 393 has proven to be far too large, which has a diminishing-returns effect on the quality of a study.

For the fifth item, if we are to allow that statistical significance is the “guard against proclaiming results that do not reflect an underlying reality, but only sampling-error happenstance” (which mere scientists who are not statisticians or philosophers of science think it means), then it is still important to understand that reflecting a real difference between two populations (treatments, networks, etc.) does not necessarily mean that difference is worth developing into a product, policy, or scientific theory.

Furthermore, statistical significance itself may not be statistically significant, meaning that a large change in levels of statistical significance can actually be the result of not-statistically-significant changes in the underlying process. The arbitrariness of the

standard thresholds (target levels) encourages dismissing medically, socially or technologically useful solutions when results at a 0.052 level may be of practical significance. This shows that relying on a single study (a single P value) is irrational. Furthermore, Tversky and Kahneman [17] have shown that even findings of statistical significance suffer from regression to the mean, and should be taken with a grain of salt. (This may explain the “decline effect” which is a growing concern unearthed by statistical meta-analysis [18, 19]. In short, the decline effect is the apparent *drastic* decline in the statistical significance of clinical-trial results after the first few studies. This has been noted, but not yet explained, and has caused some degree of jubilation in some “alternative medicine” circles since it suggests that the validity of the scientific method, and the credibility of the FDA and similar agencies should be further questioned.)

Before moving on to the problems with statistical significance testing in Artificial Neural Networks and related fields, we present a review of the basic ideas:

The upper limit for the probability of a type-I error (rejecting the null hypothesis when it is true) is set as the *significance level*, denoted by α . “We can reject the null hypothesis at a specified level of significance α only if the P -value from the sample is less than or equal to α ” [20, p. 493] because this says that the probability of observing a sample such as the one we have observed by pure chance is less than the *tolerable* probability of using such a sample to reject the null hypothesis when we should not. In other words, the significance level α is the type-I-error probability which we can tolerate. If this is the case, we say the sample is *statistically significant at the α -level*.

To express the same thing in terms of confidence intervals, we can reject the null hypothesis at the α significance level *if and only if* a $(1-\alpha)$ confidence interval does not include the null-hypothesis value [Ibid.]

Statistical significance—quite different from everyday significance—has to do with the likelihood that observed data reflect the realities of the underlying system. There is always the possibility that when a population is sampled, the observed distribution is due to pure chance (innocent sampling error beyond the control of proper design practices), and not a reflection of the underlying facts, structures, or trends. Measures of statistical significance provide a sense of how likely this is to be the case. Such measures are sometimes more meaningful if a target significance level, based on convention and the importance of rigor for the problem at hand, have been determined prior to data processing. In that case, there are two interpretations of α [2]:

If the experimenter has no *a priori* expectation, bias, belief, or hypothesis about the plausibility of a particular discrepancy, then “one begins to be slightly suspicious of

a discrepancy at the 0.20 level, somewhat convinced of its reality at the 0.05 level, and fairly confident of it at the 0.01 level. [2]

If the experimenter has any *a priori* expectation, bias, belief, or hypothesis about the underlying system (which is most often the case), this must affect her/his attitude and the choice of target significance level. “If the alternative hypothesis [were] plausible *a priori*, the experimenter would feel much more confident of a result significant at the 0.05 level than if it seemed to contradict all previous experience.” [2]

On the other hand, the idea of statistical significance and the use of certain conventional significance levels can, and frequently is, over-used to the extent that they can be less informative than simply stating the probability of observing a discrepancy as large as that observed (or larger) by pure chance. This is because common practice has made certain levels of statistical significance conventional in certain disciplines. Among these are 0.01, 0.05, and 0.10. “The statement that a particular deviation is ‘not significant at the 0.05 level’ is sometimes found to mean, on closer examination, that the actual probability is 0.06” but the *impression* that it must, then, be something close to or worse than 0.10 was made by unjustified attachment to conventional levels of significance [2].

D.2.2 Misuse of Statistical Techniques in Model Evaluation

The misuse and misinterpretation of statistics and statistical techniques is not limited to science and medicine. Engineering in general, and model selection and evaluation in the various related branches of Computer Engineering, Software Engineering, and Computational Intelligence also suffer from the same problems.

Studies of NN literature and other computational fields have shown multiple, persistent shortcomings in the quality of experimental designs and statistical analyses. Prechelt found that out of 190 articles from the mid-‘90s, only about two-thirds employed a real or realistic problem, less than one tenth employ more than one problem, and a third do not feature any comparison with another algorithm [21]. (Two of these three mistakes are avoided in the present study, and the third can be addressed with an extension of the work done to one or more of the remaining teacher-model paradigms.)

Salzberg, in addition to addressing issues of correct statistical techniques and interpretation, brings up a cautionary warning that computational fields have an additional concern: Commonly used, shared data repositories can lead to cases of

apparent statistical significance simply by statistical accident due to multiplicity in much the same way as in a single study with many comparisons.

The list goes on. More important than pointing fingers or cataloguing errors is to seek solutions.

D.2.3 What Can Be Done?

The following recommendations have been compiled [21, 22, 23, 72] to aid in the execution of high-quality scientific processes.

- Use statistical design as an ancillary to good, thoroughly evaluated scientific experimental design that is likely to give meaningful results, not just many data.
- Check assumptions and requirements.
- Select relevant procedures.
- Suspend judgment when appropriate.
- Use randomization testing and other controls¹³³
- Make accurate interpretations, and include relevant caveats.
- Keep up on the techniques and debates relevant to issues of scientific method and validity.
- Investigate Bayesian methods.

And specifically for Computational Intelligence and similar fields,

- Use separate data for design, training, tuning, and performance estimation, or use appropriate forms of cross-validation, resampling techniques, and the like if data are not numerous enough.
- Use real-world problems, not synthetic data.
- Examine more than one problem domain.

¹³³ One example is the control structure used in the present research: a completely random “prestructured” network design was created, trained, and tested in the same way as the fully connected and genuinely prestructured networks. Another example is from Salzberg [12], due to Cohen and Jensen: “For each trial, the data set is copied and class labels are replaced with random class labels. Then an algorithm is used to find the most accurate classifier it can, using the same methodology that is used with the original data. Any estimate of accuracy greater than random for the copied data reflects the bias in the methodology, and this reference distribution can then be used to adjust the estimates on the real data.” (p. 322)

- Execute multiple runs (Monte Carlo, etc.) as determined by statistical-power requirements.
- Compare against an alternative algorithm.
- Report not only the best performance, but the mean, standard deviation, statistical significance (with a scientifically informed, practical interpretation, not just blind *P*-value jockeying), and confidence intervals.

D.3 Statistics, Critical Thinking, Science, and Social Responsibility

Statistics is also an essential component of critical thinking and social responsibility. Critical thinking without an understanding of Statistics is incomplete because all decisions about the practical world involve reaching conclusions from a limited sampling of the possible situations. Statistics without critical thinking means setting oneself up to be taken advantage of, as demonstrated by the entirety of the classic work, *How To Lie With Statistics* [24], and the more recent *Seeing Through Statistics* [25], the title of which is especially illuminating: We need Statistics to see the world as it really is (for example, case study 14.1 [25]), but we often need to see *through* statistics in order not to be fooled by their presentation (for example, case study 23.1, Ibid.).

This makes it all the more important for any researcher to employ statistical techniques, justify experimental decisions, and properly interpret the results. Even in research into MUSIC (capitalized according to Elliott's distinctions¹³⁴), the responsibility of the researcher to follow the principles of scientific honesty have consequences for the reliability, defensibility and public reception and perception of Science.

This connection exists because the development (and proper use) of Statistics is intimately connected to certain findings of Cognitive Psychology, Formal Logic, Neuroscience and Evolutionary Biology. These are the cognitive illusions, logical

¹³⁴ Elliott made distinctions among “MUSIC, Music, and music” in his book on music education: “MUSIC is a diverse human practice consisting in many different musical practices or Musics. Each and every musical practice (or Music) involves the two corresponding and mutually reinforcing activities of music making and music listening. [...] The word music (lowercase) refers to the audible sound events, works, or listenables that eventuate from the efforts of musical practitioners in the contexts of particular practices.” [26]
In a similar vein, the names of fields of study are capitalized throughout this dissertation to distinguish them from specific instances of the use of their techniques. Hence “Neural Networks” refers to the field of study, whereas “neural network” means a particular network constructed in practice. Similarly, “Statistics” refers to the field, while “statistics” is the plural of a statistic, a particular figure or finding.

fallacies, biases¹³⁵ and shortcuts (*heuristics* [27; 28, p. 7; 35, p. 244]) all of which once served to increase the survival chances of our forebears, for example, by building in automatic preferences for type-I errors in predator-prey interactions¹³⁶.

¹³⁵ Anchoring[27, p. 14–16], confirmation bias, base-rate fallacy, *post hoc* fallacy, conjunction effect, framing effect, availability heuristic, clustering illusion & pareidolia, Hawthorne effect [7], stereotyping, models of mental economy, failing to recognize regression to the mean, fallacy of composition and division (as applied to complex systems with emergent properties), expectation (experimenter's) bias, distinction bias, wishful thinking, halo effect, outgroup-homogeneity bias (common in perceptions of music across cultures, such as when all African music is considered the same, but fine distinctions are made between Celtic and Scandinavian folk music), preference for anecdotes over data, semiattached figures [24: cleans 80% better ... than what?], and various memory errors [29, pp. 14–15]. For all the unreferenced entries, see [28, 30, 31, 32, 33, 34, and especially 27], as well as below.

The representativeness heuristic has to do with the extent a cause and an effect, or a process and an event, resemble each other [27, p. 4]. Things that are thus similar to each other are thought to be causally related though this is not necessarily the case. Even “experienced research psychologists” make the representativeness mistake [27, pp. 7, 9].

Regression to the mean is a (statistical, mechanical, and psychological) fact of life that is commonly misinterpreted [27, pp. 10–11]. In their performance over time in even an activity for which they are trained and specialized, all agents (athletes, pilots, teachers, students, and even machinery) perform at varying degrees of success around a personal norm, sometimes higher and sometimes lower. This implies that extremely high performance is almost always followed by lower performance, and extremely weak performance is almost always improved upon simply by natural variation, not necessarily due to reward, punishment, motivation, jinxes, praise or criticism. (In order to be able to test whether reward, punishment, etc. had an effect rather than chance variation, all factors contributing to chance variation must be somehow accounted for and controlled for in a test design.) In fact, to attribute changes in performance following outliers to the reward or punishment is to commit the *post hoc* fallacy.

The *post hoc, ergo propter hoc* fallacy is the idea that chronological order necessarily indicates causation: after this, therefore because of this.

The availability heuristic is what we use when whatever one can readily bring to mind is considered most likely [27, p. 11].

¹³⁶ An anthropoid could have been prey to large felines such as tigers, for example, and would do well to interpret *any* rustling in the bushes as a dangerous predator (tiger) rather than a small non-threatening animal (rodent). With the null hypothesis as “no threat,” a type-I error means assuming a rodent was a tiger (i.e., assuming tiger every time) and prematurely leaving the area, which in the case of a feeding or mating session, would have led to a small but acceptable loss of survival advantage. A type-II error, however, would be assuming a tiger was a rodent, which would have led to almost certain, immediate loss of life. (This sentence is not accompanied by a citation because it is the author's own interpretation arrived at upon studying cognitive psychology, statistics, logic, history of medicine and science, and evolutionary biology since 2004, initially for teaching in the University Studies program, and then for personal/scientific and research purposes.

This idea that cognitive biases evolved out of necessity and conferred evolutionary advantage is an evolutionary-psychological conjecture [35, pp. 13–14]. However, as conjectures go, it is a reasonable one. Not only does it fit the evidence of the ubiquity and usefulness [30, p. 10] of these biases, but it also enjoys circumstantial support from psychological experiments on monkeys [36] and from cognitive science. In the former case, carefully designed studies have shown that prejudice, once socially and evolutionarily advantageous, has remained ingrained in various primates, including us to this day. For the latter, in a summary of cognitive-science discoveries for aiding teaching critical

The following is an example of how the base-rate fallacy is related to type-I and type-II errors and confusion matrices:

Suppose 1000 job applicants are to take a 95% accurate drug test. Since the underlying reality of the population is that 4% actually use drugs, 40 do, and 960 do not.

Given the 95% accuracy of this fictional test (most real tests of this type have asymmetric accuracy), we can expect (in the probabilistic sense) 38 true positives (hits) and 2 false negatives (type-II errors) among the 40. We can also expect 912 true negatives and 48 false positives (type-I errors).

An individual taking the test and receiving a positive is interested in a different probability: that of a *false* positive given that one *tested* positive: $\frac{48}{48+36} = 56\%$, which is the ratio of positives for non-users to all positives.

To declare that the test is 95% accurate, hence we should be 95% confident that this person uses drugs would be to partake in the base-rate fallacy. However, we are only 56% confident that a person with a positive result actually used the drugs tested for.

Table 17: Confusion Matrix for the Base-Rate Fallacy: This type of 2-by-2 table should also be used when judging the accuracy of claims that have frequencies of occurrence as evidence. (See Table 16.)

CONFUSION MATRIX (1000)	USER (40)	NON-USER (960)
TESTED POSITIVE (86)	True Positive (38)	False Positive (48)
TESTED NEGATIVE (914)	False Negative (2)	True Negative (912)

Hence, the “base” in “base rate” refers to what is called the “prior” in Bayes’ Theorem and Bayesian analysis. Other common fallacies (that are in fact a necessary part of being a thinking organism) include *post hoc, ergo propter hoc*, meaning “after this,

thinking, van Gelder explains how critical thinking is the opposite of the type of thinking emphasized by evolution [37, p. 45]. In somewhat less detail, the idea is also mentioned in [33, p. 4]. A different but related interpretation is given in [34, p.52] in terms of “warring clans.”)

In the interest of the scientific view espoused here, it is necessary to note that no single study (no matter how well designed—the monkey study had layers of self-checking modifications) should be taken to mean a scientific fact has been established. Similarly, no conjecture, no matter how reasonable or logical, should be taken as more than a conjecture until it is transformed into a testable hypothesis and then sufficiently and reliably tested. So, it is with such a caveat that these notions of the evolutionary desirability of certain biases, shortcuts, mental preferences, and even prejudices are discussed here.

therefore because of this” [38, 39]. Occurrences following other occurrences in time are not necessarily caused by the earlier occurrences, but it appears that way to the pattern-recognizing engine in the survivor in each of us¹³⁷. In combination with confirmation bias¹³⁸ and the Hawthorne effect¹³⁹, the *post hoc* fallacy has been used in support of many controversial claims that go against well-established scientific facts *and* lack a plausible mechanism for their alleged operation.

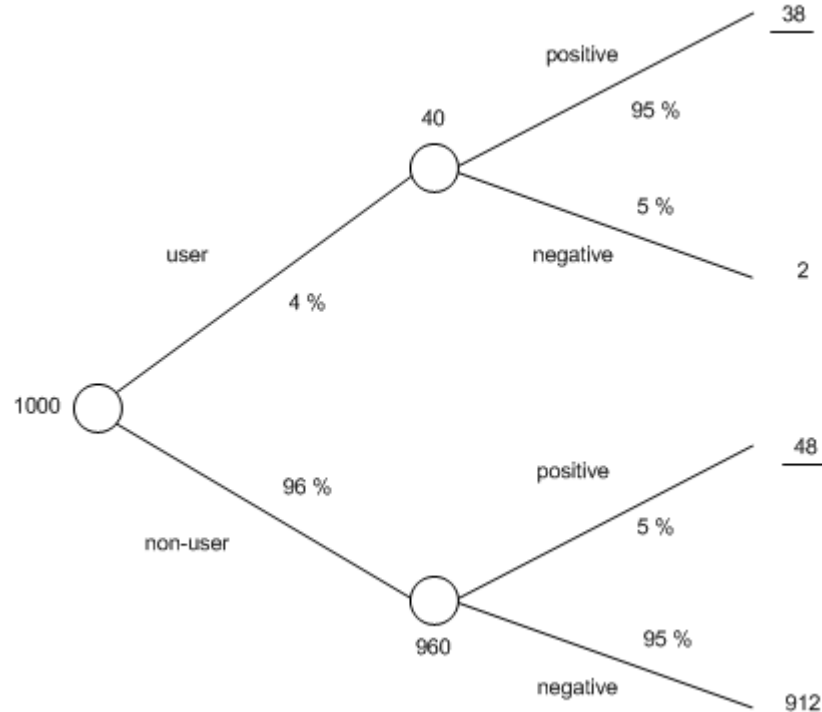


Figure 68: Probability tree with base rate (prior)

¹³⁷ As evolutionist Dawkins has pointed out, this is because *every single one of our direct ancestors* survived to reproduce, and passed at least some survivor genes to us.

¹³⁸ This is exemplified by such frustrations as “the phone only rings when I’m in the shower.”

¹³⁹ One version of the Hawthorne effect says that out of a mixed bag of results, we tend to report the better results to people—such as care-providers—who treat us well, and the worse results to those who treat us with cold professionalism

Table 18: A Version of the Confusion Matrix for Evaluating Extraordinary Claims with no Plausible Mechanism (or any claim, for that matter): Note that the only number reported (in the actual incident) was the seemingly impressive number of successes in finding water when dowsing. This is compared with the ordinary and common event of not going dowsing and not discovering underground water. The remaining cells are not typically considered.

	Gone Dowsing (unspecified number of times)	Not Gone Dowsing (most of the time)
Success	An impressive number, such as 10	Unknown
Failure	Not reported	Most of the time

Another example relevant to the present research is taken from an appendix to a master's thesis in music in which a famous Brazilian musician and music educator makes the fallacious claim that "Brazilian music does not have a clave" [40].

The complete statement in response to the question "Is partido alto similar in any way to clave in Afro-Cuban music?" is as follows:

I sincerely don't think so. Just because clave is also a two[-]bar pattern[,] it does not mean that they are similar. Maracatu (from Recife) is also a two[-]bar pattern. Do you think Maracatu has anything to do with clave? Brazilian and Cuban rhythms, as you know, are two completely different things, like bananas and watermelons. I don't know one Cuban musician that understands Brazilian rhythms 100% or vice versa. BRAZILIAN MUSIC DOES NOT HAVE A CLAVE. [40, p. 38; emphasis in the original]

This statement by Da Fonseca (whose work was consulted as part of the present research) is a veritable gallery of fallacies: circular reasoning, the argument from ignorance, false dichotomy, and hasty generalization.

Setting aside the fact that the interviewer's question was whether there was *any* similarity between *partido-alto* and Afro-Cuban clave, not whether they were the same thing, consider first the example of *maracatú*. Da Fonseca takes it for granted that *maracatú* has nothing to do with clave. This is a matter of the definition of clave, and his reasoning is circular. If one defines clave as solely Cuban, then the argument is moot; all things Brazilian are devoid of clave. However, such a definition is neither useful nor practically accurate. Clave, as a concept, is practically and theoretically more relevant to any discussion (other than that of particular patterns) than clave the pattern, and

maracatú certainly follows its own pattern of higher and lower offbeatness around the phrase cycle.

The statement that Brazilian and Cuban rhythms are like two different fruit is a false dichotomy. Bananas and watermelons have plenty in common. If the issue under consideration is biological reproductive function for plants or vitamin content for animal nutrition, the two are very similar and highly related. However, they are not the same fruit, and neither are they completely unrelated; there is a gray area. The same is true of Brazilian and Cuban rhythms: They follow the same type of conceptual rules of rhythmic harmony, even if some details may vary.

Subsequently, a hasty generalization is employed: Da Fonseca knows no Cuban musicians who fully understand samba; therefore no Cuban musicians can fully understand samba. The inductive leap is unnecessary and offensive. This is also an argument from ignorance, as lack of evidence of Cubans who fully understand Brazilian music is seen sufficient to claim it is not possible, and furthermore, that this is because the two musics have nothing in common in terms of their temporal aspect, which is not a justified conclusion.

As we have seen, the necessary cognitive biases of human thought and the consequent logical fallacies arise in every aspect of life, from science and pseudo-science to music theory and culture.

To further explore this connection between Science, Statistics, Evolution, and Cognitive Psychology, we continue with a brief survey of primate history:

Around 50 million years ago, in the Eocene epoch [41, p. 56; 42, p. 13], “the first anthropoids (a subgroup of primates consisting of the monkeys and apes) appear in the fossil record” [41, p. 56] of the then-tropical forests of North Africa’s Mediterranean coast [42, p. 7]. “[T]his date really marks the root of the human line” [41, p. 56] as the wet-nosed primates (lemurs, etc.) split from the dry-nosed primates (our line) [42, p. 13, 122–124].

This is relevant to issues of human decision-making because the dry-nosed primates are capable of the types and levels of thought already superior to that needed to judge risk and benefit in potential predator interactions [42, pp. 134–135]. The uncertainties necessitated by the incompleteness of the fossil record [41, pp. 31, 122; 42, pp. 215, 217, 228] may be hiding errors in the details of the argument, but do not obscure its basic point.

Regardless of which group of primates we take as the starting point for this discussion, ***the time period is of the order of hundreds of thousands or even millions of years.***

- anthropoids (the so-called higher primates),
- hominids (the so-called great apes, who are thought to have split off from the other apes 6–8 million years ago [41, p. 153; 42, p. 187]),
- humans (those hominids that are of the genus *Homo*, “2.4 to 1.5 million years ago” [41, p. 153; 42, p. 197]),
- recent humans (*Homo erectus* and the like, from about one million years ago [41, p. 153] until a few hundred thousand years ago [42, p. 198], or
- modern humans (since around 600 thousand years ago [41, p. 153]),
Homo heidelbergensis, *Homo sapiens*, and *Homo neanderthalensis*¹⁴⁰,

A particular cutoff is beyond the scope of this appendix, as well as an artificial, purely taxonomic notion [43, pp. 106–107]. The main point is that ***a very long time, and a large portion of our history, has passed without the benefit of widespread critical inquiry, widespread communications, and Science***, and under the yoke of necessary but harmful biases and fallacies, no matter how we slice who “we” are or what intelligence is. This is discussed here because of its relation to the argument below about the evolutionary basis of the difficulty in the spread and acceptance of modern critical scientific modes of thinking, and the role of Science, as well as science communication in social responsibility.

It is not until partway through the Miocene (until about 5 or 6 million years ago) that we suspect *consciousness* appeared, as correlated with the emergence of *Homo sapiens* (but not yet the current subspecies) [41, pp. 59–61, 89, 158; 42, pp. 185–192]. Even those 5 million years of *Homo sapiens sapiens* history have mostly passed without the benefit of writing. Writing, which began approximately 5400 years ago, makes possible sustained sharing of thoughts and discoveries. Thus we can estimate that *at least* about 47 million years [42, p. 7] of anthropoid history lacked the benefit of thinking-about-thinking, and all but about 5400 years of the 5-million-year history of humans was completely devoid of writing [44, first page of map 5]. Furthermore, even after the invention of writing, 75 to 90% of the populations of most societies remained unable to read until the Industrial Revolution¹⁴¹. During all that time, the survival advantages

¹⁴⁰ which may indeed be *sapiens* and not a separate species, but certainly coexisted with *Homo sapiens sapiens* (the subspecies) quite recently and for about ten thousand years) [42, p. 199–201].

¹⁴¹ Even after the invention of writing, it is estimated here that 75 to 90% of most societies were illiterate until the Industrial Revolution—this estimate was conjunctively arrived at by considering qualitative and quantitative evidence in [46, 47, 48]. Furthermore, according to various UN and OECD reports, women continued to have literacy rates below 25% in 19 countries (mostly in Asia and Africa) as of the mid-1990s [44, pp. 74–75, 120].

conferred by our overactive pattern-recognition ability¹⁴² unchecked by critical thinking [30, p. 10] served our forebears well, and did so right up until the relatively rapid developments of agriculture, writing, the industrial revolution and telecommunications technology.

Once these “sudden” developments came about, however, the mental habits (hardwired biases) that once had survival value became impediments to reasoning. The conditions and vicissitudes of the modern (agricultural/industrial/scientific/Internet) age require a very different set of heuristics *and* a freedom from cognitive biases.

Evolution may give us these improved mental attitudes and abilities over another billion years, but most likely will not do so, for the same reasons as the persistence of pain [29, pp. 393–396]. The role of Science, then, is *not only* to fuel the ever-accelerating development of technology and medicine (as is commonly thought) and to uncover the mysteries of the universe (which is, in an ideal sense, its core purpose), but also to make up for the difference in speed between our genetic evolution¹⁴³ and our extremely rapid social and technological evolution.

Accordingly, part of the role of scientists in society, is to “prevent harm” [45, p. 216]. A beautifully concise and powerful analogy for this role is given by Olson:

“Their job is to question *everything*. This is what scientists do for a living: they are trained *not* to take bait. When you give a scientist a paper to read, [...], he or she will question the premise, question the assumption, demand to see the data, demand that you cite your sources This is why the phrase “Scientists agree...” actually means something. [...] They view themselves as the designated drivers of [society]. While everyone [else] gets drunk on entertainment, the scientist maintains a certain level of sobriety, always keeping an eye on the facts.” [1, pp. 92–93; emphasis added]

Furthermore, “the public has to *demand* that scientists protect them from misinformation” [45, p. 217]. The misinformation scientists must protect the public from is not only in the areas of nanotechnology or stem cells, but also in areas as diverse as psychology, biology, cosmology, medicine, pseudo-science, history, renewable

¹⁴² For an example, consider the Gregory masks: [28, p. 33–35] or online: <http://www.richardgregory.org/experiments/>.

¹⁴³ Biological evolution does not occur fast enough to adapt our brains to the new challenges of having gone from being the prey to being the universal predator. Social and technological changes occur at a rate that is orders of magnitude greater. This is true irrespective of whether Darwin’s *gradualism* or Gould’s and Eldredge’s *punctuated equilibria* wins the ongoing debate on the rate of evolution.

energy, world affairs, and the statistics people encounter every day on the news. Much of this misinformation gets generated by scientists through misunderstanding, misapplication, and misinterpretation of statistics. This is doubly troubling because scientists, technologists, and academics have the additional responsibility of watching out for and correcting society's errors, not adding to them. Our social and technological evolution has far surpassed the rate of our biological (especially cognitive) evolution, and the difference has to be made up for by those who specialize in thinking critically.

Not only does biological evolution operate on a much slower rate than technological development and social change, but being a chance process with no conscious purpose (illustrated elegantly by “the drunkard’s walk” in [49, Fig. 21, p. 148]), evolution has in fact “hard-wired” [37, p. 45] these “features” (Ibid.) into our brains (Ibid.):

...the mind has intrinsic tendencies toward illusion, distortion, and error. To some extent, these are just features of the ‘hard-wired’ neural equipment we inherited through the accidental process of evolution¹⁴⁴. To some extent, they are the results of common patterns of growth and adaptation—the way our brains develop as we grow up on a planet such as Earth. To some extent, they also are ‘nurtured,’ that is, inculcated by our societies and cultures. Yet, whatever their origin, they are universal and ineradicable features of our cognitive machinery, usually operating quite invisibly to corrupt our thinking and contaminate our beliefs. These tendencies are known generically as ‘cognitive biases and blindspots.’

Furthermore, van Gelder explains, “a majority of people ... do not have a general grasp of the notion of evidence” [37, p. 42] and that “[h]umans are not naturally critical [because] evolution does not waste effort making things better than they need to be, and [*Homo sapiens*] evolved to be just logical enough to survive” [37, p. 42]. (For a more detailed, yet relatively brief, discussion of our built-in faulty reasoning and the necessary shortage of critical thinking, see [35, pp. 242–9].)

An interesting historical connection between the musical aspect of this dissertation (*samba carioca*) and ignorance about science and medicine lies in the story of how Rio’s poor moved up the hills into the *favelas* where samba came into its own: During the highly violent Vaccine Revolt of 1904, when Dr. Oswaldo Cruz mandated smallpox vaccination, this was perceived by the poor people of the city as an effort to kill off the poor [50, pp. 178–180]. The required vaccinations were part of a “serious programme of reconstruction and sanitation” [50, p. 178] which included demolishing

¹⁴⁴ This accidental process is only one step in the process of evolution, which is a deterministic process with naturally occurring random elements, *not* a random process.

the poor neighborhoods in the center of the city, the residents of which moved up into the hills (*morros*) [50, p. 178], ultimately giving rise to today's "samba schools," one from each hill.

The Brazilian revolt in 1904 differs from today's anti-vaccine movement in the US in terms of its methods (no gun-fights in the US thus far over vaccines), but quite similar in terms of the scientific ignorance at its root. The current vaccine controversy is based on a single discredited and retracted article that features conflicts of interest, fraudulent data, and the *post hoc* fallacy [51–55].

It is this role of Science—making up for the difference in rates between biological and societal–technological evolution in order to allow humans to be critical thinkers by building on the discoveries of past geniuses (or simply imaginative, hard-working individuals) and using the checks and balances of modern scientific principles—that researchers, authors, social scientists, and physical scientists in the following list have been increasingly vocal about¹⁴⁵:

- Kahneman & Tversky [27]
- Gilovich: *How We Know What Isn't So: The Fallibility of Human Reason in Everyday Life* [30, specifically pp. 186, 193]
- Kida: *Don't Believe Everything You Think: The 6 Basic Mistakes We Make in Thinking* [31]
- Schick, Jr. and Vaughan: *How To Think About Weird Things: Critical Thinking for a New Age* [32]
- Anderson: *The Complete Thinker: A Handbook of Techniques for Creative and Critical Problem Solving* [56]
- Philips: *The Undercover Philosopher: A Guide to Detecting Shams, Lies, and Delusions* [57]
- Bausell: *Snake Oil Science: The Truth about Complementary and Alternative Medicine* [7]
- Randy Olson: *Don't Be Such a Scientist: Talking Substance in an Age of Style*, very pointedly on p. 5: "the entire fate of humanity" [1]
- Ruggiero: *Beyond Feelings: A Guide to Critical Thinking* [21]

¹⁴⁵ All this is relevant to a dissertation in Electrical & Computer Engineering because a Doctor of Philosophy degree differs from a professional doctorate in the central role of theory and experimentation, which are coordinated by the scientific method, and which must be carried out in a responsible manner. In turn, Science and the scientific method (as the author has learned over the course of his PhD studies) are not separable from society and its politics, or from philosophy itself. The references [28, 30–32, 56, 57] have been used as textbooks in the *Knowledge, Rationality & Understanding* cluster (coordinated sequence of University Studies courses) that the present author taught in. Several of the others were considered or used by faculty for additional material.

- Piattelli-Palmarini: *Inevitable Illusions: How Mistakes of Reason Rule Our Minds* [28]
- Neil DeGrasse Tyson¹⁴⁶
- P. Z. Myers¹⁴⁷
- Carl Sagan: *The Demon-Haunted World: Science as a Candle in the Dark* [58]
- Richard Dawkins: *The Greatest Show on Earth: The Evidence for Evolution* [29] and *Climbing Mount Improbable* [43]
- Leonard Mlodinow: *The Drunkard's Walk: How Randomness Rules Our Lives* [33]

The accelerating rate of technological and social change requires conscious adaptation to the new circumstances through a system of informed practices that may be collectively called critical thinking. The five components of critical thinking are:

- Cognitive Psychology: understanding one's own thinking and natural biases,
- Philosophy: rationality and the recognition of fallacies,
- Quantitative Literacy: the ability to comprehend, interpret, use and explain geometric, numerical, symbolic, and statistical information,
- Information Literacy: the know-how to judge the quality and objectivity of information sources such as websites, blogs, TV and radio news, books and magazines,
- Cultural and Intercultural Competence: possessing an understanding of cultural differences—not merely specific to a given situation, but the wide spectrum of potential misunderstandings as well—along with a broad general knowledge.

Cognitive Psychology informs us of cognitive biases and illusions, which are default starting points and misinterpretations that are evolutionarily built into our thinking. Philosophy, specifically the study of logic, sheds light on fallacies, which are mistakes in reasoning. Quantitative literacy involves developing a working knowledge of Mathematics, the physical sciences, and as often overlooked, Statistics, which is the philosophical and mathematical framework that validates the practice of Science. This is why Statistics plays such a central role in all experimental design.

The remaining elements of critical thinking are also essential to the present work. Information literacy is the ability to evaluate sources for their accuracy, bias and rigor. This can be taken to the idealistic extreme of checking all claims independently

¹⁴⁶ For astrology and similar beliefs, see:

<http://www.haydenplanetarium.org/tyson/watch/2008/06/19/debunking-astrology> and

<http://www.haydenplanetarium.org/tyson/watch/2008/06/19/full-moon-effects>.

For “Stupid Design,” see: <http://www.youtube.com/watch?v=YGKRurORkCA>

¹⁴⁷ For example, on the issue of alternative-medicine, see <http://scienceblogs.com/insolence/>, and on general biology topics, <http://scienceblogs.com/pharyngula>.

for oneself (which is infeasible), or reduced to the more common practice of relying on the skepticism and self-check of the scientific community (precisely defined below). The necessity for relying on the latter means full scientific honesty and rigor are even more essential than previously implied.

Finally, cultural and intercultural competence (especially relevant due to the central place afforded to an etic formalization of a musical theory of rhythmic syntax for Afro-Brazilian music as part of the present research) will be briefly discussed from the point of view of scientific concern as evidenced in the academic debates of recent decades between the physical sciences and the postmodern¹⁴⁸ academia, which ironically share the same ideals of knowledge, understanding and social justice.

In addressing these debates, the present author, by virtue of belonging to both groups, can take both the position of a multiculturalist from a non-Euro/American background with a liberal-arts degree and longstanding interests in Ethnomusicology and Ethnomathematics, as well as that of a “northern¹⁴⁹”-educated PhD-seeking humanist with degrees in Mathematics–Physics and Electrical & Computer Engineering, thus, once again, seeking a balance between opposing influences.

The place of Science is seriously threatened in today’s world, and the accuracy of some of Science’s best accomplishments are downplayed or even denied by religious extremists, some religious moderates, some of the news media, many artists, most postmodern academics and intellectuals, and self-described environmentalist or multiculturalist liberals (those who are relativists); in other words, just about everyone [1, 59, 60]. To address this problem, there have been multiple calls in the scientific and technological community for scientists and engineers to communicate more effectively with the rest of society, from the magazine of the IEEE¹⁵⁰ and the journal *Nanoethics* [45, pp. 215–220], to the recent books *Don’t Be Such A Scientist* [1, pp. 7–11 and throughout] and *Unscientific America* [75, p. 124–5 and throughout].

It would also be valuable to communicate more effectively with one another about our work, which includes (but is not limited to) better-informed practices and clearer reporting and understanding of statistical techniques in science and technology (including their mathematical and philosophical underpinnings and quandaries).

¹⁴⁸ The term postmodern is used here as a catchall phrase that includes any rejection of rational thought and of Science (as just one way of knowing among many) [57, pp. 225–231], including those approaches associated with cognitive relativism, extreme structuralism, social constructivism [60, pp. 4–17], and poststructuralism (such as that of Bergson, who seems to have based a career and a school of philosophy on a profound misunderstanding of Einstein).

¹⁴⁹ See Appendix G.

¹⁵⁰ The Institute of Electrical and Electronics Engineers is the world’s largest and most influential social and technical organization for electrical, computer and electronics engineering.

Before wrapping up the discussion of critical thinking, then, it may help to address the danger that Science and critical thinking face from the indiscriminating, reactionary trend of postmodern cognitive relativism.

In an ambitious work entitled *Human Accomplishment: The Pursuit of Excellence in the Arts and Sciences, 800 B.C. to 1950*, Murray claims:

“Judgment is separable from opinion in matters of artistic and scientific excellence. It is possible to distinguish the important from the trivial, the fine from the coarse, the credible from the meretricious, and the elegant from the vulgar. Doing so is not a simple matter, and no single observer is infallible, but a realm of objective knowledge about excellence exists. That knowledge can be tapped systematically and arranged as data that meet scientific standards of reliability and validity.” [59, p. *xvi*; italics in the original]

This is a tough claim to support. Who determines the elegant from the vulgar? Just like the “Supreme Court” way of telling pornography from art, when it comes down to a fine line, the decision is up to an individual’s call (or one society’s values). This way of determining the elegant from the vulgar is *not* scientific, as we have seen when Europeans first encountered Africans and described their intricate musical rhythms as primitive and vulgar.

Such a claim, then, may raise the question of whether Murray will argue that certain cultures or ethnic groups are superior to others in their propensity for great art or science. Fortunately, Murray dismisses the possibility of such an approach by the following page, recognizing that all human societies have the same potential for excellence, but that the “environments for eliciting great accomplishment” [59, p. *xvii*]—not to mention, chance—have a role in determining when and where they take place. Murray does not reconcile his two claims, but the present author will attempt reconciliation below, proposing different attitudes for accomplishments that are measurable (in the sciences and technology) and those that are not (in the arts and humanities).

However, for the moment, let us return to Murray to clarify why his work is quoted above in the first place. The conflict between the two academic worldviews is interpreted by Murray as follows:

“The disillusionment following the World Wars has since given rise to a broader intellectual rejection of the idea of progress. The idea of the Noble Savage, another fancy of the Enlightenment, has reemerged in our own time. It has become fashionable to decry modern technology. Multiculturalism, as that word is now understood, urges us to accept all cultures as equally praiseworthy. Who is to say that the achievements of

Europe, China, India, Japan, or Arabia are ‘better’ than those of Polynesia, Africa, or the Amazon? Embedded in this mindset is hostility to the idea that discriminating judgments are appropriate in assessing art and literature, or that hierarchies of value exist—hostility as well to the idea that objective truth exists.” [59, p. *xviii*]

Murray appears to equate multiculturalism and the associated relativism with the idea that all accomplishments are equal, rather than the idea that all cultures have accomplishments, or that different cultures have had different circumstances (hence needs and drives) to achieve those. The present author subscribes to a much milder form of multiculturalism (one informed by Information Theory and Statistics) that is not anti-science, but does not devalue cultural alternatives either. Likewise, defended in this appendix is a more limited form of relativism, as explained below.

At the other end of the spectrum, the subjects of Sokal & Bricmont’s controversial critique [60] are presented as advocating—and may well be—the idea that scientific or mathematical discoveries, such as the value of π or the theory of general relativity, are culturally determined, and would have turned out different had they been discovered or developed by “non-northern” (cf. Appendix G) peoples.

The present author takes the following intermediate (and interdisciplinary) stance: Although appropriate for the humanities (which involve matters of taste, such as music) relativistic approaches are not appropriate for explanations of how the universe works (at least if what is desired is anything other than either the Sceptical¹⁵¹ immobility of being confined to Plato’s cave, or an anything-goes free-for-all where scientific theories are replaced by flights of fancy and wishful thinking). Specifically in the case of the humanities, while differences exist among contributions by various societies and cultures, it is more likely that qualitative differences dominate over quantitative ones, or at the very least, it is impossible to judge quantitative differences in such multivariate endeavors.

For example, different cultures may cultivate (focus on, devote energy to, and develop) different aspects of the same art form, leading to different ideas of what constitutes excellence. Members of those cultures, then, tend to take into account *only* the aspects which they have cultivated when evaluating the accomplishments of other cultures (whose societies may have been focused on other aspects considered less important by these observers).

To put it simply, with examples from the realm of music, one culture may cultivate to a remarkable extent the musical elements of harmony, counterpoint, or

¹⁵¹ The spelling convention used here is “skeptical” for modern scientific skepticism and “sceptical” for the Sceptics of ancient Greece.

texture (all the while certain that they have not forsaken other aspects like rhythm, pitch or structure). Another culture may eschew polyphony completely, and develop a sensitivity to pitch (throughout the culture, not only in its musician class) so fine that other cultures do not consider those intermediate pitches (“microtones”—the term itself is obviously *etic*¹⁵²) to be proper pitches in and of themselves.

Yet another culture may delve just as finely into the *time* between note events, developing rhythmic nuances foreign to all other parts of the world, while neither neglecting nor thoroughly cultivating the focal points for the aforementioned cultures.

Other examples may be listed, but should not be necessary. It is equally relevant to point out that cultures are also differentiated by another aspect of their approach to excellence: Some develop elaborate formal theories which are taught in schools, both informing and reflecting the evolution of their traditions; others (with just as elaborate, but informal theories) disseminate tradition in a less precise, less elitist, aural and oral manner.

These differences imply neither that the northern approach is *better* (as implied by proponents of the outdated social-Darwinian notion that technological and scientific progress indicates overall superiority), nor that the non-northern approach is *better* (as implied by reactionary/anti-science proponents of a post-modernist flavor of multiculturalism where acupuncture is assumed superior to “Western” (northern) medicine solely by virtue of its geographical origin).

Furthermore, the assertion that both approaches and both sets of contributions are valid and valuable is not a politically correct cop-out, but reflects a necessary difference in the application of critical thinking to the arts versus the sciences: There are many ways of knowing in the arts; there are *no* certain ways of knowing in the sciences. But, there are better ways of approaching phenomena to ascertain the likelihood of truth for a given hypothesis. It is for this epistemological difference (as supported by the opposing forces of scientific objectivity—also precisely defined below—and the problem of induction [57, pp. 234–7; 73, pp. 4–17]¹⁵³) that Statistics is the science of Science, and a critical component to its execution.

To invoke the term “science of Science” necessitates that “Science” be defined. “Science” is not simply the act of pursuing such activities as chemical reactions, sequencing genomes, or validating mechanical principles. Properly pursued, Sociology,

¹⁵² The anthropological terms *emic* and *etic* refer to the points of view of insiders and outsiders, respectively.

¹⁵³ The consistency of Science, which we observe, is based on the assumption of the consistency of its observations. This is challenged in the problem of induction: Is there any way to be certain that universal physical constants are universal or constant? Did they hold the same values a few microseconds after the Big Bang as they do now? Do they hold the same values in remote areas of the universe now as they do in our galaxy (relativistic concerns of “now” notwithstanding)?

Psychology, and other social sciences constitute *Science*, whereas dishonest or careless work (even if conducted while wearing a lab coat) does not. (Fost, in fact, gives a description of how it would be possible, or may even be common, to practice the physical sciences at a university or research institute and not *be a scientist* [61, pp. 208–9].) The distinction requires explaining what is meant by “the scientific method.”

D.4 The Scientific Method, Past and Present

In the diagnosis of disease, Hippocrates introduced the elements of the scientific method. He urged careful and meticulous observation: “Leave nothing to chance. Overlook nothing. Combine contradictory observations. Allow yourself enough time.” Before the invention of the thermometer, he charted the temperature curves of many diseases. He recommended that physicians be able to tell, from present symptoms alone, the probable past and future course of each illness. He stressed honesty. He was willing to admit the limitations of the physician’s knowledge. [58, p. 8]

“There was no such thing as the Scientific Revolution, and this is a book about it.” [62, p. 1] So starts Shapin’s *The Scientific Revolution*. The UCSD sociologist continues to give an excellent distillation of the reasons neither the scientific revolution nor the scientific method actually exist as such [62, pp. 3–4, 6]. Nonetheless, as even Shapin is quick to point out [62, p. 5], something did take place in early modern Europe that gave rise to the modern, broader scientific method(s) whose consequences the world enjoys not only in the realm of intellectual curiosity, but also in terms of medicine and technology.

Developed in Renaissance Europe both from the remains of ancient Greek science¹⁵⁴ and as an intellectual revolt against its hold on the centers of learning [62, pp. 5, 10, 43, 54, 65–69, 78, 80; 30], the early scientific method was a methodical combination of three ideas. These were logic: the method of resolution and composition from the medieval era, now known as analysis and synthesis; experimentation: the value of repetition, as practiced by the artist-engineers of the Renaissance; and theory: a combination of the philosophical, mathematical, and early scientific contributions of the ancient Greeks and Indians as preserved by Arab scholars [64]. Yet, at the same time as it was based on the western European interpretation of

¹⁵⁴ Ancient Greek science was initially fed by the mathematics, philosophy and natural observations of the Babylonians, Chaldeans, Egyptians and Indians [65], and later preserved and augmented by the secular-scientific (“peripatetic”) period of the Arab and Islamic society (roughly 750–1400) [66, pp. 38–132]

ancient Greek philosophy, the scientific method of the Renaissance was also a rebellion against the methods and ideas of the established Aristotelian learning of the late-Medieval and early-Renaissance periods [63, pp. 238–240, 250].

This particular set of ideas and practices were united into the cyclical four-step process commonly referred to as the scientific method: observation, hypothesis-forming, experimentation, and either theory-formation upon verification, or re-framing the problem with a new hypothesis and returning to experimentation. This is at least in part due to Bacon, who was the chief proponent of the inductive scientific method [63].

There are three problems with this definition. One is that it may not be necessary to strictly follow this model. In exploratory research (See Section 2.1), experimentation may be done without specifying a hypothesis because the purpose of the research is to *generate* hypotheses rather than evaluate them.

Secondly, it appears that scientists, at least some of the time, actually follow a rather different sequence where observation follows experimentation. Many academics (including scientists) have expressed concerns about the idea of there being a scientific method at all (for example, [57, p. 245]), and others argue strongly that the inductive method is not it (for example [57, pp. 218–9]).

One of the ways in which we can reconcile these important views and still allow that such a thing as *Science* exists may be to explain the process in terms of the hypothetico-deductive method:

From a general hypothesis and particular statements of initial conditions, a particular predictive statement is deduced. The statements of initial conditions, at least for the time, are accepted as true; the hypothesis is the statement whose truth is at issue. By observation we determine whether the predictive statement turned out to be true. If the predictive consequence is false, the hypothesis is disconfirmed. If observation reveals that the predictive statement is true, we say that the hypothesis is confirmed to some extent. A hypothesis is not, of course, conclusively proved by any one or more positively confirming instances, but it may become highly confirmed. A hypothesis that is sufficiently confirmed is accepted, at least tentatively.

Good examples for the preceding notion are given in [57, p. 247]. While the hypothetico-deductive method is different from Bacon’s inductive scientific method, it does involve inductive inference, though at a different point along the process.

As described in [2], the hypothetico-deductive method holds a “lens” or “window” up to the “true state of nature” (which is not directly observable, by definition—see ancient Greek sceptics and modern Neuroscience). This lens or window is a “designed experiment D_j , which collects data related to the true state of nature, but

adds noise. This data may be combined with any previously available data, and in combination with a set of consequences previously derived for a hypothesis H_i (which was derived through deduction, as will be expounded below). The data and the consequences of the hypothesis at hand are applied to inductive inference, possibly yielding a modified hypothesis H_{i+1} .

The original hypothesis, however it came about, had previously led to two developments: the design of the experiment D_j , and the use of deduction to arrive at the consequences of that hypothesis.

Next, if necessary, hypothesis H_{i+1} replaces H_i , leading to possible changes in the experiment (and the collection of more data), as well as the deduction of the consequences of the new hypothesis. These two are subjected to further inductive inference, and the cycle ideally continues until a satisfactory hypothesis is reached.

It is in the approach to any satisfactory hypothesis that further disagreement is found in the philosophy of science (the initial disagreement being that hypotheses are not created by induction from observations, but rather that observations are made because hypotheses are put forth, possibly through culturally conditioned insight).

Objections to either model abound, and both the acclaimed philosopher Salmon [38] and the acclaimed mathematician Jaynes [75] have put forth Bayesian solutions to the joint crisis of Kuhn/Popper/Hume. The details of reconciling the difficulties in the philosophical underpinnings of various approaches to and interpretations of Science are beyond the scope of even this far-reaching appendix, but one may refer to Salmon for his proposal to establish the validity of the scientific method by reconciling the hypothetico-deductive method, Popper, and other challenges by using the Bayesian framework [38, pp. 114–129]. To put it very briefly and simple-mindedly, the Bayesian framework enjoys two types of support, one from its role in solving the base-rate problem in everyday estimation, and the other from its successful track record of solving technological problems, though the latter does not get around Hume's objection that induction may not be used to justify induction.

It is relevant, however, to point out, that the music-theoretic portion of the research described in this dissertation came about entirely as prescribed (or described) by the *inductive* scientific method. First, in October 2001, the observation was made that the so-called 3-2 “bossa clave” was qualitatively different from the 3-2 *son* clave. Measurements of inter-onset intervals were made—a form of data collection—and a hypothesis was formed as a result. Over the next several months, this hypothesis was applied to the *cáscaras*, the *rumba* clave, and patterns from samba, salsa, and *maracatú*. Eventually, upon comparisons with patterns from many clave-based styles, the initial hypothesis was found to be lacking, and through more observation, a new hypothesis was formed. By the time the dissertation was ready for defense, the clave-direction hypothesis had gone through about half a dozen iterations, including some major

changes to reflect the observations (including changes regarding its scope). This is a scientific approach to change direction, and shows that inductive scientific inquiry does occur.

The scientific method, whatever it really is, has changed since the Renaissance and the Enlightenment in a sense that is perhaps more important than these debates. It has progressed toward better insuring scientific honesty, reliability and objectivity. It is these contemporary principles of scientific thinking that concern us here more than the arguments about hypothetico-deductive versus inductive ultimate approaches and whether either is followed precisely in practice. These fundamental principles are Accuracy, Objectivity, Skepticism, and Open-mindedness [35, p. 17]. They, in turn, are supported by other experimental/statistical techniques and principles listed subsequently.

D.4.1 Accuracy

Accuracy refers to “gathering and evaluating information in as careful, precise, and error-free a manner as possible” [Ibid.]¹⁵⁵.

D.4.2 Objectivity

Objectivity is “obtaining and evaluating such information in a manner as free from bias as possible” [Ibid.]. ‘Bias’ in this case refers to the cognitive biases that are natural to human thinking and judgment, such as confirmation bias, Hawthorne effect¹⁵⁶, selection bias, etc. (A more complete list was given in Section E.5.)

¹⁵⁵ In the general scientific sense, accuracy includes precision as a subset. In terms of the properties of sampling distributions (such as experimental results), however, accuracy and precision are independent. Accuracy relates to validity (degree of correctly reflecting the underlying reality), whereas precision relates to reliability (degree of consistent repeatability). A popular depiction shows a dartboard with a tight cluster of darts far off the center. The tightness indicates consistency (precision); the distance indicates lack of accuracy. A more typical example in science education is making the wrong measurement with overly ambitious precision: the wrong result, to many significant figures. In popular culture, precision without accuracy is exemplified by the Hardcore Punk ethic “Be strong; be wrong.”

¹⁵⁶ The Hawthorne effect has multiple interpretations (<http://www.cs.unc.edu/~stotts/204/nohawth.html> and <http://www.psy.gla.ac.uk/~steve/hawth.html#warn>), and the interpretation intended here is not the original (<http://www.library.hbs.edu/hc/hawthorne/09.html>), which has been disputed, but that of Bausell, as related to selective memory. Bausell states that patients will unintentionally emphasize the positive parts of their experiences following a treatment (when reporting on that treatment) if they perceive the caregiver to be caring, kind and concerned. That is, for a time period during which they perceived both improvement and deterioration of their condition, they will focus on the periods of improvement when reporting to a sympathetic, caring physician, but will report deterioration as well when reporting to a physician who comes across as cold and impersonal [7].

D.4.3 Skepticism

Skepticism is defined precisely as the willingness to accept findings “only after they have been verified over and over” [Ibid.] by many different, reliable researchers.

D.4.4 Open-mindedness

Open-mindedness means to *not* resist changing one’s own views—even those that are strongly held—in the face of evidence that they are inaccurate [Ibid.].

D.4.5 Additional Concerns, Techniques, and Principles

To these four principles, we can add several others. For conjectures and hypotheses, we have plausibility, falsifiability, confirmation, and corroboration. As hypotheses move toward the status of theory, we are again concerned with falsifiability, confirmation, and corroboration, and also with transferability. Additional statistical tools for rigorous pursuit of these goals include randomization, blocking, and bootstrapping, as well as the experimental techniques of double-blinding (to avoid confounding variables due to biases that are external to the problem addressed by the research), and factorial design (including and controlling for confounding variables that are inherent to the problem addressed by the research).

The reason for including critical thinking and issues of postmodernist interpretations in this dissertation is that the scientific method and the use of Statistics are intimately tied in with the information-theoretic concerns of the present research: parsimony, goodness-of-fit, and Occam’s Razor, discussed in Sections 1.2 and 2.5, which in turn are applications of Skepticism, which is diametrically opposed to unscientific thinking and emotional reasoning.

The balance between constraint and entropy (cf. Section 2.5) is analogous to the balance between parsimony and goodness-of-fit. To put it simply for the purpose of a quick-and-dirty analogy, parsimony is the preference for simpler models while goodness-of-fit typically favors more complex models. Simpler models correspond to greater constraint; more complex models correspond to better fit. Both concerns are addressed by a historical progression of the notion of parsimony:

- Aristotle on parsimony: “Nature operates in the shortest way possible.” [67]
- John Duns Scotus & William of Ockham: Given two good explanations, the one that is less complex ought to be selected when no other information is available.
- Laplace’s principle of insufficient reason: Do not select a more elaborate probability distribution than you have information to support [68, pp. 4–14].

- Newton’s first rule of reasoning: “We are to admit no more causes of natural things than such as are both true and sufficient to explain the appearances. To this purpose the philosophers say that Nature does nothing in vain, and more is in vain when less will serve; for Nature is pleased with simplicity, and affects not the pomp of superfluous causes” [69].
- Einstein: “A theory is the more impressive the greater the simplicity of its premises is, the more different kinds of things it relates, and the more extended is its area of applicability” [70].
- Lendaris/Stanley: Given two networks that can successfully learn a mapping, the one with the smaller performance subset will generalize better to conditions previously not experienced [71].

At this point, it is necessary to note that while these statements are very similar in their gist, some are intended for explanations of nature (those of Aristotle, Scotus and Ockham, Newton, and Einstein), one is intended for statistical assumptions applicable to all phenomena (that of Laplace), and one is intended specifically for a technological application (that of Lendaris and Stanley for an application based on an imitation of nature). The present research has sought and confirmed evidence that the principle of parsimony is likely to apply to technological artifacts as it is expected to apply to nature. (A Popperian objection to this statement may be that this is not sufficient or even acceptable because evidence was not sought to falsify the applicability of parsimony to the technological case, but this is not true: The experimental setup was neither set to confirm nor falsify, but to investigate both outcomes fairly; i.e., every effort was made to design the best standard network (null hypothesis, status quo) in an apples-to-apples comparison with the best prestructured network for each of the three “finalist” prestructuring techniques.

The argument has been made (in the Statistics and model-evaluation literature) that techniques developed for the social sciences may not apply to the physical sciences or technology. If so, parsimony, a basic principle for the study of nature (physical science), may not apply to other domains. On the other hand, parsimony is *not a natural law* that can be verified, but a probabilistic guide for inquiry, and thus potentially applicable to technology.

There is an urban myth that comes to the defense of applying principles of parsimony to guide engineering design and management decision-making, although the true story behind this urban myth demonstrates a key difference between explanations of nature and technical problem-solving.

It is falsely rumored that because regular ball-point pens did not function well in zero-gravity, NASA spent millions of dollars to develop a “space pen” while the

Russians simply used pencils. Regardless of the factual nature of the details of the story, this is a great argument for seeking parsimony in design. However, it turns out that American astronauts also initially used pencils, but a scientist/inventor voluntarily developed the special pen because broken pencil lead floating unhindered by air resistance proved to be a hazard.

A key difference between science and technology is revealed by this example: In technology, there can be multiple good solutions with their respective trade-offs. In the sciences, however, there can only be one correct explanation. For instance, Maxwell's equations or the laws of thermodynamics either hold or do not. It is not the case that they hold for Russians and not for Americans. (For those wishing to evoke the problem of induction and humankind's confinement to a very small slice of space-time, the preceding statements can be taken to apply only to this limited slice without detracting from the argument.) Hence it may be expected that the applicability of parsimony is limited (but not nonexistent) in the technological case. Nonetheless, neural networks, being both stochastic systems and imitations of nature, have been shown in this dissertation to benefit from an application of parsimony to their design.

References for Appendix D

- [1] Olson, R., *Don't Be Such a Scientist: Talking Substance in an Age of Style*, Washington, DC: Island Press, 2009
- [2] Box, G. E. P., Hunter, W. G., and Hunter, J. S., *Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building*, New York, NY: John Wiley & Sons, p. 109, 1978.
- [3] McClave, J. T., and Sincich, T., *Statistics*, Upper Saddle River, NJ: Prentice Hall, p. 94, 2003.
- [4] Miller, J., "Statistical significance testing—a panacea for software technology experiments?" *The Journal of Systems and Software*, 73, pp. 183–192, 2004.
- [5] Cohen, J., "A Power Primer," *Psychological Bulletin*, Vol. 112, No. 1, 1992, pp. 155–159; p. 158.
- [6] Carver, R. P., "The Case Against Statistical Significance Testing," *Harvard Educational Review*, Vol. 48, No. 3, 1978, pp. 378–99.
- [7] Bausell, R. B., *Snake Oil Science: The Truth About Complementary and Alternative Medicine*, Oxford, 2007, Chapters 1 & 2.
- [8] Freedman, K. B., Back, S., and Bernstein, J., "Sample size and statistical power of randomised, controlled trials in orthopaedics," *J Bone Joint Surg Br*. Vol. 83-B, No. 3, 2001, p. 397.
- [9] Ioannidis, J. P. A., "Why Most Published Research Findings Are False," *PLoS Medicine*, Vol. 22, No. 8, 2005, p. e124.
- [10] Siegfried, T., "Odds Are. It's Wrong: Science fails to face the shortcomings of statistics," *Science News – The Magazine of the Society for Science & the Public*, Vol. 177, No. 7, 2010, p. 26.
- [11] Ziliak, S. T., and McCloskey, D. N., *The Cult of Statistical Significance: How the Standard Error Costs Us Jobs, Justice, and Lives*, Ann Arbor, MI: The University of Michigan Press, 2008.
- [12] Salzberg, S. L., "On Comparing Classifiers: Pitfalls to Avoid and a Recommended Approach," *Data Mining and Knowledge Discovery*, Vol. 1, 1997, pp. 317–328.
- [13] Freedman, K. B., and Bernstein, J., "Current Concepts Review - Sample Size and Statistical Power in Clinical Orthopaedic Research," *J Bone Joint Surg Am.*, No. 81, 1999, pp. 1454–60.

- [14] Gelman, A., and Stern, H., “The Difference Between “Significant” and “Not Significant” is not Itself Statistically Significant,” *The American Statistician*, Vol. 60, No. 4, 2006, pp. 328–331.
- [15] Berger, V. W., PhD, and Hsieh, G., PhD, “Rethinking Statistics: Basing Efficacy Alpha Levels on Safety Data in Randomized Trials,” *Israeli Journal of Emergency Medicine*, Vol. 5, No. 3, 2005, pp. 55–60.
- [16] Lenth, R. V., “Some Practical Guidelines for Effective Sample-Size Determination,” (submitted 2000, revised 2001), Department of Statistics, Obermann Center for Advanced Studies, University of Iowa.
- [17] Tversky, A., and Kahneman, D., “Belief in the law of small numbers,” In D. Kahneman, P. Slovic, and A. Tversky (eds.), *Judgment Under Uncertainty: Heuristics and Biases*, Cambridge: Cambridge University Press, 1971, pp. 3–20.
- [18] Lehrer, J., “The Truth Wears Off: Is there something wrong with the scientific method?,” *The New Yorker Annals of Science*, 2010.
Available: www.newyorker.com/reporting/2010/12/13/101213fa_fact_lehrer
- [19] Rabin, Y., “Fundamental Limitation on Applicability of Statistical Methods to Study of Living Organisms and Other Complex Systems,” *Journal of Statistical Physics*, Vol. 144, 2011, pp. 213–216.
- [20] Albright, S. C., Winston, W. L., and Zappe, C., *Managerial Statistics*, Monterey, CA: Brooks/Cole Publishing Company, 2001.
- [21] Prechelt, L., “A quantitative study of experimental evaluations of neural network algorithms: Current research practice,” *Neural Networks*, 9, 1996.
- [22] Flexer, A., “Statistical Evaluation of Neural Network Experiments: Minimum Requirements and Current Practice,” In R. Trappl, ed., *Cybernetics and Systems '96: Proc. 13th European Meeting on Cybernetics and Systems Res.*, 1996, pp. 1005–1008.
- [23] Golden, R. M., “Statistical tests for comparing possibly misspecified and non-nested models,” *Journal of Mathematical Psychology*, 44, 2000, pp. 153–170.
- [24] Huff, D., *How to Lie with Statistics* (47th printing), New York, NY: W. W. Norton & Company, 1982.
- [25] Utts, J. M., *Teaching Seeing Through Statistics: An Instructor's Resource Manual*, Belmont, CA: Duxbury Press, Wadsworth Publishing Company, 1996.
- [26] Elliott, D. J., *MUSIC MATTERS: A New Philosophy of Music Education*, New York/Oxford: Oxford University Press, 1995, pp. 44–45.

- [27] Kahneman, D., Slovic, P., and Tversky, A. (eds.), *Judgment under uncertainty: Heuristics and biases*, Cambridge, UK: Cambridge University Press, 1982.
- [28] Piattelli-Palmarini, M., *Inevitable Illusions: How Mistakes of Reason Rule Our Minds*, Wiley, 1996, p.7.
- [29] Dawkins, R., *The Greatest Show on Earth: The Evidence for Evolution*, New York, NY: Free Press, 2009.
- [30] Gilovich, T., *How We Know What Isn't So: The Fallibility of Human Reason in Everyday Life*, New York, NY: The Free Press, 1991.
- [31] Kida, T., *Don't Believe Everything You Think: The 6 Basic Mistakes We Make in Thinking*, Amherst, NY: Prometheus Books, 2006.
- [32] Schick, Jr., T. and Vaughn, L., *How to Think About Weird Things: Critical Thinking for a New Age*, Mountain View, CA: Mayfield Publishing Company, 1999.
- [33] Mlodinow, L., *The Drunkard's Walk: How Randomness Rules Our Lives*, New York, NY: Vintage Books, 2009.
- [34] Daley, J., "What You Don't Know Can Kill You," *Discover: Science, Technology, and The Future*, Waukesha, WI: Kalmbach Publishing Co., July/August 2011, pp. 50–57, 88.
- [35] Baron, R. A., and Kalsher, M. J., *Essentials of Psychology*, Needham, MA: Allyn & Bacon, A Pearson Education Company, 2002.
- [36] Grewal, D., "The Evolution of Prejudice," *Scientific American*, April 5, 2011. Available: www.scientificamerican.com/article.cfm?id=evolution-of-prejudice
- [37] Teaching Critical Thinking: Some Lessons From Cognitive Science, T. van Gelder, *College Teaching* (Heldref Publications), Vol. 53, No. 1, 2005, pp. 41–46.
- [38] Ruggiero, V. R., *Beyond Feelings: A Guide to Critical Thinking*, Boston, MA: McGraw Hill, 2004.
- [39] Mooney, C., and Kirshenbaum, S., *Unscientific America: How Scientific Illiteracy Threatens Our Future*, New York, NY: Basic Books, 2009.
- [40] Wilson, D., Partido Alto: *Rhythmic Foundation Analysis of Aquarela Do Brasil*, master's thesis, Youngstown State University, 2010, p. 38.

- [41] Silvertown, J., (ed.), *99% Ape: How Evolution Adds Up*, London, UK: Natural History Museum, 2008.
- [42] Tudge, C., and Young, J., *The Link: Uncovering Our Earliest Ancestor*, New York, NY: Little, Brown and Company, 2009.
- [43] Dawkins, R., *Climbing Mount Improbable*, New York, NY: W W Norton & Company, 1997.
- [44] Haywood, J., Catchpole, B., Hall, S., and Barratt, E. *Atlas of World History*, Black, J., Chaudhuri, K. M., Cunliffe, B., Fagan, B. M., and Spence, J.E. (eds.), Oxfordshire, UK: Andromeda, 1997.
- [45] Evans, N. G., “Speak No Evil: Scientists, Responsibility, and the Public Understanding of Science,” *NanoEthics*, Vol. 4, No. 3, 2010, pp. 215–220.
- [46] Hunt, L., Martin, T. R., Rosenwein, B. H., Hsia, R.P., and Smith, B. G., *The Making of the West: Peoples and Cultures (A Concise History), Volume II: Since 1340*, Boston, MA: Bedford/St. Martin’s, 2003, p.586.
- [47] Ehrman, B. D., *Misquoting Jesus: The Story Behind Who Changed the Bible and Why*, New York, NY: HarperOne, 2005, pp. 32–39.
- [48] Goldschmidt Jr., A., *A Concise History of the Middle East* (Seventh Edition), Cambridge, MA: Westview Press, 2002, p. 219.
- [49] Gould, S. J., *Full House: The Spread of Excellence from Plato to Darwin (Uncorrected Proof/Special Advance Reader’s Edition)*, New York, NY: Harmony Books, 1996.
- [50] Castro, R., *Rio de Janeiro: Carnival under Fire*, [translated by John Gledson from the original *Carnaval no fogo*, Companhia das Letras, Brazil], London, UK: Bloomsbury Publishing Plc., 2004.
- [51] Nicoll, A., Elliman, D., and Ross, E., “MMR vaccination and autism 1998: Déjà vu—pertussis and brain damage 1974?,” *BMJ* (formerly *British Medical Journal*), Vol. 316, No. 7133, 1998, p. 715.
- [52] Elliman, D., and Bedford, H., “MMR vaccine: the continuing saga: Current concerns are idiosyncratic: most reviews have confirmed the vaccine's safety,” *BMJ* (formerly *British Medical Journal*), Vol. 322, No. 7280, 2001, p. 183.
- [53] Deer, B., “Secrets of the MMR scare: How the case against the MMR vaccine was fixed,” *BMJ* (formerly *British Medical Journal*), Vol. 342, No. c5347, 2011.

- [54] Deer, B., "The Lancet Scandal"
Available: briandeer.com/mmr-lancet.htm
- [55] The United Kingdom Department of Health, "MMR: *the facts*,"
Available:
www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/documents/digitalasset/dh_105189.pdf
- [56] Anderson, B. F., *The Complete Thinker: A Handbook of Techniques for Creative and Critical Problem Solving*, Englewood Cliffs, NJ: Prentice-Hall, Inc., 1980.
- [57] Philips, M., *The Undercover Philosopher: A Guide to Detecting Shams, Lies, and Delusions*, Oxford, UK: Oneworld Publications, 2009.
- [58] Sagan, C., *The Demon-Haunted World: Science as a Candle in the Dark*, New York, NY: Ballantine Books, 1996.
- [59] Murray, C., *Human Accomplishment: The Pursuit of Excellence in the Arts and Sciences, 800 B.C. to 1950*, New York, NY: HarperCollins Publishers, 2003.
- [60] Sokal, A., and Bricmont, J., *Fashionable Nonsense: Postmodern Intellectuals' Abuse of Science*, New York, NY: Picador, 1998.
- [61] Fost, J., *If Not God, Then What?: Neuroscience, Aesthetics, and the Origins of the Transcendent*, Clearhead Studios, 2007.
- [62] Shapin, S., *The Scientific Revolution*, Chicago and London: The University of Chicago Press, 1996.
- [63] Hall, M. B., *The Scientific Renaissance: 1450–1630*, New York, NY: Dover Publications, Inc., 1994.
- [64] Rice, E. F., Jr., and Grafton, A., *The Foundations of Early Modern Europe, 1460–1559*, New York, NY: W. W. Norton & Company, 1994.
- [65] Ronan, C. A., *Science: Its History and Development among the World's Cultures*, New York, NY: Facts on File Publications, 1983, pp. 31–44, 61.
- [66] Hoyt, E. P., *Arab Science: Discoveries and Contributions*, Nashville and New York: Thomas Nelson Inc., Publishers, 1975.
- [67] <http://math.ucr.edu/home/baez/physics/General/occam.html>

- [68] Kapur, J. N. and Kesavan, H. K., *Entropy Optimization Principles with Applications*, Academic Press, San Diego, 1992.
- [69] Newton, I. Sir (1687), *Sir Isaac Newton's MATHEMATICAL PRINCIPLES of Natural Philosophy and his System of the World*, Translated into English by Andrew Motte in 1729. The translations revised, and supplied with an historical and explanatory appendix, by Florian Cajori, Berkeley: University of California Press, 1962, p. 398.
- [70] Hawking, S. (ed), *A Stubbornly Persistent Illusion: The Essential Scientific Works of Albert Einstein*, Philadelphia, PA: Running Press, 2007, p. 353.
- [71] Lendaris, G. G., and Stanley, G. L., "Self-Organization: Meaning and Means," Inf. Syst. Sci.; Proc. 2nd Congress on Inf. Syst. Sci., Baltimore, MD: Spartan Books, 1965.
- [72] Vurkaç, M., "Some Problems & Solutions in the Experimental Science of Technology: The Proper Use and Reporting of Statistics in Computational Intelligence, with an experimental design from Computational Ethnomusicology," Portland State University Systems Science Seminar Series, February 25th, 2011. Available: <https://sas.illuminate.com/site/external/jwsdetect/playback.jnlp?psid=2011-02-25.1246.M.6D598EC9FA3A19B4F7B15EDAC00EB6.vcr&sid=2009262> and <http://www.pdx.edu/sysc/seminar#Winter2011> (slides)
- [73] Salmon, W. C., *The Foundations of Scientific Inference*, Pittsburgh, PA: University of Pittsburgh Press, 1979.
- [74] Gould, S. J., *Full House: The Spread of Excellence from Plato to Darwin (Uncorrected Proof/Special Advance Reader's Edition)*, New York, NY: Harmony Books, 1996.
- [75] Jaynes, E. T., *Probability Theory: The Logic of Science*, Cambridge, UK: Cambridge University Press, 2003.

APPENDIX E: The Data-Acquisition Process

An engineer just makes things work. But the artist asks profound, provocative questions: What feelings does this evoke? How does this relate to the whole? What does it mean? We need to look at the entire picture...Division is dangerous. (Hiroshi Ishii, associate director of the Media Lab at MIT [1])

The process of extracting, understanding, codifying, and using musical knowledge as gathered through musical practice is generally of a different nature than the scientific process. The extraction of any type of musical knowledge is not readily implemented as a controlled experiment. Because the information sought is of a subjective nature, or at least culturally constructed and context-dependent, the process of obtaining it is iterative and experiential.

A number of musicologists and ethnomusicologists have explained how process can fit into the scientific method. In his Ethnomusicology Master's thesis on clave at Tufts University under David Locke, Lehmann describes the role active musicianship plays in the gathering of data and the formulation of theories.

The present author's approach to data collection has precursors in Lehmann's approach, as well as his predecessor Washburne's approach to studying the clave of salsa music [2, p. 2]:

"I have determined to wear, simultaneously, the hats of researcher (fieldworker) and the actively performing musician. Thus, playing salsa music has become my primary methodological tool . . . the process of data collection is often conducted `on the bandstand` . . . by the use of trial and error, imitation of insider's (sic) behavior, and informal `hanging out` sessions with other performers. [. . .] My data[-]acquisition process as both a performer and researcher are then one and the same. Both require an acute awareness of one's surroundings, an ease in adapting to new situations, and an ability to organize, analyse, and interpret a large body of information." (p. 2)

Similarly, the present author has worn the hats of researcher, performing musician, and instructor intensively from 2001 to 2009. During this time, his primary focus has been to learn about the particular sense of clave direction found in Brazilian music. This was accomplished through several avenues, such as:

- practicing, performing with, and occasionally directing one of the nation's leading *blocos* (large Brazilian-style percussion-and-dance ensemble) of *samba batucada* 1997 through 2009;
- teaching weekly samba-percussion classes, thus learning from the questions and ideas of, and difficulties faced by the students;
- starting and leading a samba-based cover band featuring the author's re-arrangements of classic and modern-popular *axé*, *ijexá*, and samba songs;
- attending California Brazil Camp (seven times, 2001–2011) to study samba, *samba-reggae*, *samba-afro*, *baião*, and *Candomblé* with Jorge Alabê (of Mocidade and Ôba-Ôba, and *alabê* of Casa Branca, Ketu Nation), Gamo da Paz (of Ballet Folclórico do Brasil, Ballet Folclórico da Bahia, and *gamo* of Gantois, Cantuá, Ketu Nation), Carlinhos Pandeiro de Ouro (of Mangueira and the movie *Black Orpheus*), Marcio Peeter & Wagner Profeta Santos (of Ilê Aiyê), Anderson Pandeiro (of Mangueira), Mark Lamson (of Bata Ketu fame), Bruno Moraes & Alex Rangel (of Mocidade), Airto Moreira (of Weather Report, Return To Forever, Miles Davis, Dizzy Gillespie and Stan Getz), Boca Rum, Curtis Pierre (of Casa Samba), Michael Spiro (of Ella Fitzgerald, David Byrne, Changuito, Giovanni Hidalgo, Gilberto Gil, Eddie Palmieri, Clark Terry, McCoy Tyner, Bobby McFerrin, sambamasterclass.com and Bata Ketu), Jorge Martins (percussion director of Maracatu Nação Estrela Brilhante), Dudu Fuentes (co-director of Bangalafumenga), Vanderlei Pereira, and other renowned performers;
- applying and testing out *partido-alto* and clave-related ideas to four rock bands with styles ranging from trip hop to sludge metal, as well as contrasting the Brazilian tradition with Afro-Cuban, Ghanaian, Japanese, Balkan, and Renaissance traditions, which he has studied and performed; and
- workshops with other Brazilian musicians, such as Claudinho Sorriso (samba song), and Cuban master Jesús Alfonso.

While all of these experiences have provided opportunities to test out ideas, the bulk of the author's musical research has been in *samba carioca* (Rio-style samba), through workshops, performances, lessons, research sessions, and practices with Jorge Alabê, Michael Spiro, Mark Lamson, Derek Reith, Brian Davis and Andrew Hartzell.

Throughout this process, the author has maintained critical thinking and skepticism, leading to several revisions, broadenings, and narrowings of hypotheses about the functioning of clave direction (through *partido-alto*) in samba. This iterative process of data-collection-through-music-making (and learning) is echoed by Lehmann

in his discussion on Nzewi's remarks about the role of both objective and subjective approaches to the research and analysis of music [2, p. 2]:

Nzewi [3] remarks on [...] a researcher's 'em-etic approach' who is 'capable of objective self-assessment': 'Cognition should be predicated on total facultative and emotive submission to the affective object of perception. We, therefore, need to be seeking a researcher/analyst of African music who is at the confluence of objective observation (by virtue of correct and applicable scientific training) and subjective perception (by virtue of enculturation or genuine penetration of the cultural intentions and creative manifestations of a cultural phenomenon).'

The present author enjoys precisely such a synthesis of scientific training and "genuine penetration of the cultural intentions and creative manifestations" of *samba carioca*.

However thorough these studies and analyses were, at the onset of the present work, the present author naively thought that the classification of the 65536 rhythm patterns described in section 4.2 into three clave categories (generally called 3-2, 2-3, and neutral—the fourth category was not considered at the time) would be a near-trivial task for any expert samba percussionist. This task had to be completed in order to generate training, test, and validation sets. Since the 65536 vectors make up the complete input space, the various data sets would all be drawn from these vectors, and randomly ordering them would be necessary for the I/O space to be reliably spanned by each of the three sets. To this end, the rhythm patterns were ordered using a high-quality pseudo-random-number generator, and the classification process was begun. Patterns were classified according to their interactions with four standard parts: the "tricky" *samba de roda* pattern (1), *partido alto* (2), standard¹⁵⁷ *tercera*, or *third surdo* (3), and the pattern known as *teco-teleco* (4).

During classification, it was discovered that the qualitative observations readily made for many of the patterns arose from two factors. One was that members of a given clave-direction category were found to be of that category for different (rhythmic) reasons. The other factor, which became apparent over a much longer period of time during the classification process, was that membership in a given clave-direction category could be a matter of degree. These observations were recorded as comments on the classification process itself as well as on the vectors to which they pertained, in

¹⁵⁷ By standard, what is meant is the most common of the standard *tercera* parts used by *Grupo Especial* samba schools, exemplified by G.R.E.S. Mocidade Independente de Padre Miguel.

an additional field of the data file. Some of these comments expressed the degree of conviction as to the assigned clave direction, while other comments had to do with styles of Afro-Latin music in which these patterns were typically encountered or were most likely to occur¹⁵⁸. Still other comments expressed speculation as to the dependency of the final clave-direction assignment on tempo and pitch (which are not part of the attack-point representation used here).

Several thousand patterns were classified. The task was tedious: It involved staring at lines of zeros and ones while mentally (though without difficulty) converting them to rhythmic patterns, and if necessary, playing these patterns against some fundamental traditional pattern like *partido alto*, *tercera* or the *bossa clave*¹⁵⁹. Within a few weeks, it was clear that the classification of arbitrarily chosen pitchless, durationless, non-traditional patterns into the three clave categories was not at all a trivial task, and subtleties of membership in these clave categories were discovered and noted. In addition to the three clave-direction categories, 885 patterns out of 6326 were notated as “neither” or “none” (eventually dubbed “incoherent”). This had the potential to be a new discovery, at least in terms of being explicitly stated; the category *clave-incoherent* is not part of the lore of clave-based musics to the extent experienced by the researcher. However, this notion of incoherence was soon to be verified by the expert percussionists M. Spiro and M. Lamson, as recounted below.

Considering that mistakes could have been made due to exhaustion after several hundred patterns, audits were conducted. Indeed, upon review of the first several thousand category assignments, disagreements (with the current pass) were found, and a second round of classification was done. This round, in addition to checking for mistakes, served the purpose of some degree of reclassification informed by discoveries from the first round regarding the complications and subtleties of clave direction. The earliest-encountered patterns did not have the benefit of the perspective of having seen or heard dozens or hundreds of others. A screen capture of several patterns and the associated classification and annotations from this early stage are given in Figure 69.

¹⁵⁸ As the intent to derive a “unified theory of clave” was replaced by the drive to provide engineering research with a reliable data source, this concern was replaced by the question of how a pattern would be perceived in the context of *samba carioca* regardless of where it was most likely to be found.

¹⁵⁹ The researcher can play three rhythmic parts simultaneously without the use of instruments, as long as one is a metronomic pulse. The pulse and the related count are kept via a specific pattern of stepping developed by the samba community in Portland, Oregon for clave-phrasing and time-keeping purposes. In addition, a traditional pattern like *bossa clave* or *partido alto* can be clapped while the pattern in question is “sung” in syllables.

Though the concept of a no-clave category (meaning neither clave direction, eventually dubbed “incoherent”) was being formed, a systematic way of identifying this category had not yet been developed. The process of observing, listening, and playing the thousands of patterns for a second time helped crystallize the ideas behind how the neutral and incoherent categories differed. Data files were saved under separate names, and another pass with specific schemata¹⁶⁰ in mind for detecting patterns in the incoherent category was executed at the end of May 2007 with full agreement in all previously classified 3-2 and 2-3 patterns.

E.1 Data Acquisition: Golden Ears

Still, the process was daunting. In February 2008, three of the leading clave and samba experts in the United States visited Portland to conduct workshops and offer private lessons. Research sessions were arranged by the author (as private lessons) with each of the master musicians. These sessions were actually a form of “golden-ear experiment,” named after the listening tests audio- and recording-equipment manufacturers perform with audiophiles (individuals with highly trained and tuned ears, who can pick out slight distortions inaudible to the average person). In this case, the golden-ears were not audiophiles, but *clavephiles* or *sambaphiles*. One of the experts is a lifetime samba and *Candomblé* teacher, performer and ordained member of the clergy who won the first Brazilian national award for instrumental proficiency on the *repique* (the lead drum in samba). The two other experts who took part in this portion of the study are among the few non-Cubans ordained in Cuba for performing the religious rites of the Afro-Cuban *batú* drums, which can be traced back to some of the same African roots as Brazilian *Candomblé*, which in turn is the foundation of samba music. They are also internationally recognized experts of samba performance.

These experts were presented with a selection of the most challenging patterns (from the point of view of clave-classification) identified during the first two passes. The patterns were presented along with “control” patterns—well-known traditional patterns that any *sambista* would agree on the orientation of. They were performed by the author who was separated from the golden-ear by a screen, and accompanied by another expert musician playing a traditional part with a well-known clave-orientation, and without variation. The patterns were presented in multiple ways (phase shifts and inversions) for reliability¹⁶¹.

¹⁶⁰ Sections with a strong correlation to membership in a clave-direction class.

¹⁶¹ The third expert did not require the method of physically performing the patterns because he saw the binary-vector representation used by the author, and stated that he also conceptualized the rhythms as binary sequences. He proceeded to write down the corresponding clave direction next to each

0011	0111	0011	1001	N	"all possibilities for relative off-beatness, so CAUTION	
0111	1001	1000	1110	3	By a little	
1000	0011	0011	0110	3		
0101	0000	1110	0000	3		
0111	0111	0011	1111	3	By a little	
1101	1111	1000	1010	N		
1010	1110	1011	1110	2	By a little	
0110	1100	1110	1011	3	Just barely	
1000	1000	1000	1000	0	Absolutely	
0100	0111	1000	1110	3	By a little	
0010	1101	0000	1011	2		
1001	0011	1100	1100	3		
1011	1001	1000	1111	2	Just barely	
1101	0010	1110	0111	3		
0000	1101	0000	0001	2		
0011	0000	0011	1000	2	By a little	
1000	0001	1010	1100	3		
0010	1111	1100	1100	2	Just barely	
1110	1101	1101	1110	2		
1101	1100	1011	0101	3		
0010	0101	0011	1110	2	By a little	
0010	1011	1100	1001	2	2-3 Rumba all the way; Partido OK	
0000	0011	1100	1011	2		
0001	0000	0111	0010	2	By a little	
0111	0001	1101	1101	3	By a little	
1000	1000	1110	1110	3	By a little	
1000	0101	0110	1110	2		
1101	1010	0100	0001	3	Just barely	
1101	1101	0111	1100	3	Bossa+Rumba work	
0100	0011	1011	1000	3		
0110	1100	1001	0100	3	Just barely	
0111	1001	1111	0011	2	Just barely	
0001	0001	1010	0010	3	Rumba	
0111	0111	0000	1000	3	Mostly Son	
0010	0010	1101	1111	2	Just barely	
0101	0001	0110	0111	3	Just barely	
1000	1001	1000	1110	2	"3-2 Rumba almost, but more 2-3 Partido and Bossa	

Figure 69: A screen capture showing early classification of and comments on the randomly ordered patterns. (Because the final four categories were not explicitly recognized at this time, the coding is not the same one used later on during the bulk of the research.)

The results were not encouraging. For the highly obscure patterns, the three experts agreed on 40 out of 73 cases. For example, the pattern 0010 | 1000 | 1010 | 1001 was classified as being “OK over 3-2” by only two of the experts, whereas only one

binary vector. This gave a different type of result than the first two trials, where if the expert’s answer was ‘no’ (or ‘crossed’), the only way to judge the expert’s idea of the clave direction was to compare their response with the response to the shifted presentation of the same rhythm.

agreed with Brazilian practice for 1010 | 1010 | 1010 | 1001. (This pattern was taken from [6].) Another example is 0111 | 1111 | 0000 | 1110, on which a different selection of two experts agreed, and more importantly, two showed disagreement between their respective responses to this pattern and the following very similar pattern in the same direction (giving rise to questions of intra-rater reliability and supporting the cause of the present research): 0111 | 1110 | 1001 | 0001.

However, a number of very useful insights were gathered from the three experts' responses, especially in their comments, which were written down by the note-takers:

- 1) Different experts might rate the patterns' membership in a clave direction with different degrees of strictness.
- 2) It is important for the designer of the study to specify the cultural context because clave-direction is interpreted differently in different countries (or regions), and in musics originating from different time periods (modern music vs. traditional music). It is similarly necessary for the expert to specify which cultural context they are considering in classifying a particular pattern because experts with knowledge of more than one tradition classify the patterns according to the tradition the pattern in question is closest to, even when asked only to consider one tradition (such as *samba carioca*).
- 3) Further support was found for the fourth clave-direction category (incoherent).

These were all important discoveries for the author, who had not, in the course of learning and performing clave-based musics from Brazil, Cuba and Ghana, encountered any patterns that clearly fell so far outside the concept of clave-direction because, by design, clave-based musics follow clave rather closely most of the time.

Along with the fourth clave-direction category, the most important discovery was the possibility of classifications based on musical context as well as teaching context, with the latter resulting in the different levels of strictness codified into the later stages of the research.

The loose/lenient¹⁶² classification can be interpreted as relating to the use of transient patterns associated with non-repeating musical devices such as a solo, a lick, a *subida* (a type of samba entrance consisting of a crescendo of eighth-note triplets, or a variation of such), or a variation. In contrast, the strict classification has to do with steady-state rhythmic elements like *riff*, *solí*, *vamp*, *ostinato*, *head*, or *theme*. Certain patterns

¹⁶² The classification is “loose” if the person doing the classification is “lenient.”

that may not be acceptable as, say, 3-2, when repeated (used as *ostinati*) may still be acceptable as 3-2 if used once in passing, whether as part of an introduction, or in modulating from one rhythm to another, or as part of an improvised solo.

The strict-vs.-lenient denomination was also overloaded with another contextual wrinkle: In addition to the sense that a pattern can be looped and still fit into the stated clave direction, *strict* was chosen to mean that it can be applied to any of the primary samba percussion instruments (*surdo*, *caixa*, *agogô*, *tamborim*¹⁶³). While admittedly problematic, this conflation of meanings was deemed necessary in order to avoid a foray into various concerns of context, which could defeat the purpose of using the attack-point abstraction and require more detail than can be reasonably addressed in a pioneering study. Therefore, the extreme ends of the context spectrum were selected as the strict and lenient classification schemes.

The loose/lenient classification scheme, then, in addition to the loosening of the requirement of repeatability (loopability/vampability), also acquired the acceptance of being applicable to only a proper subset of the four primary samba instruments.

As a result, a pattern that would be considered, say, 3-2, when played in passing by a certain instrument such as the *agogô* (but not by the *caixa*) would be considered 3-2 in the loose sense, but not in the strict sense. Only patterns that were allowed to be repeated on any of the four primary instruments and still be considered 3-2 would be categorized as 3-2 under the strict sense.

E.2 Data Acquisition: Classification

A third classification run was then performed by the researcher. This pass incorporated three new ideas: strict and loose classification contexts, the fourth clave-direction category (incoherent), and the strict limitation of the cultural context to Rio-style samba (*samba carioca*). Hence, concerns about how a pattern would be likely to be interpreted in music from Cuba, or even elsewhere in Brazil were set aside. Some 6000 patterns were completely reclassified in March 2008 (having erased the prior category assignments). Observations from the “strict” expert and the “lenient” expert were incorporated into the “strict” and “loose” classifications, both including the fourth clave-direction (incoherent). Furthermore, in order to avoid assigning patterns to categories to which they may only sometimes (under certain conditions or contexts)

¹⁶³ While these ideas are applicable to other instruments, including strings and wind instruments, for simplicity, these will not be considered in the present research.

belong, detailed comments were included about how certain patterns may be interpreted if played on one instrument versus another.

After this, 60 more patterns—most of which were known to the author to be common practice in samba—were added directly to the training set on the grounds that when humans learn to play the music, they mostly encounter patterns of strong and clear clave direction like these. Then in April 2008, these direct-to-training-set vectors were augmented by another 83 rhythms taken from the books *O Batuque Carioca*, *Pandeiro Brasileiro*, *Ritmos do Brasil para Bateria*, *New Ways of Brazilian Drumming*, *Brazilian Rhythms for Drumset*, *Brazilian Percussion Manual*, and one relevant pattern from the Uruguayan-music book *El Toque de Candombe*¹⁶⁴. These were checked for redundancy with what was already in the set of randomly selected vectors so as to avoid duplication. A subset of these are shown in Figure 55.

0	0	0	0	1	1	0	1	0	0	0	0	1	1	0	1	3	O Batuque Carioca, p. 20, #4
0	0	0	0	1	0	1	0	0	0	0	0	0	1	0	1	1	O Batuque Carioca, p. 20, #6
0	0	0	0	1	0	0	1	0	0	0	0	0	1	0	1	1	O Batuque Carioca, p. 20, #7
0	0	0	0	1	0	1	0	0	0	0	1	1	1	0	1	1	O Batuque Carioca, p. 20, #8
0	0	0	0	0	0	1	0	0	0	1	0	0	0	1	0	1	O Batuque Carioca, p. 20, #9
0	0	0	0	1	0	1	0	0	0	1	0	0	1	0	1	1	O Batuque Carioca, p. 20, #10
1	0	0	1	1	0	0	1	1	0	0	1	0	1	0	1	1	O Batuque Carioca, p. 22, Imperio Serrano (caixa)
1	0	0	1	0	0	1	0	1	0	0	1	0	1	0	0	1	O Batuque Carioca, p. 23, Uniao da Ilha/Grande Rio/Beija Flor/Mangueira (caixa)
1	0	1	1	1	0	1	0	1	0	1	0	1	0	0	1	1	O Batuque Carioca, p. 23 Salgueiro (tarol)
1	0	1	1	1	0	1	1	1	0	1	0	1	0	0	1	1	O Batuque Carioca, p. 23 Vila Isabel (tarol)
0	1	0	1	1	1	1	1	1	0	1	1	1	1	1	1	1	O Batuque Carioca, p. 25, repique solo 2
1	1	0	1	1	0	1	1	1	0	1	0	0	1	0	1	1	O Batuque Carioca, p. 25, repique entrada 1/2
0	1	0	1	0	1	1	0	1	0	1	0	1	1	0	1	1	O Batuque Carioca, p. 25, repique entrada 1/2
0	1	0	1	1	0	1	0	1	0	1	0	1	1	0	1	1	O Batuque Carioca, p. 25, repique entrada 1/2
0	1	0	1	1	0	1	0	1	0	1	0	1	1	0	1	1	O Batuque Carioca, p. 27, tamborim (levada 3)
1	0	1	1	1	0	1	1	1	0	1	1	0	1	1	0	1	O Batuque Carioca, p. 31, agogo 1
0	1	0	1	1	0	1	1	1	0	1	0	1	1	0	1	1	O Batuque Carioca, p. 31, agogo 2
1	0	1	1	0	1	1	0	1	0	1	1	1	0	1	1	1	O Batuque Carioca, p. 31, agogo 3
1	0	0	1	1	0	1	1	1	0	1	1	1	0	0	1	1	O Batuque Carioca, p. 31, agogo 6
1	0	0	0	1	0	1	0	1	0	0	1	1	0	0	1	1	O Batuque Carioca, p. 31, agogo 7
0	1	0	1	1	0	1	1	1	0	1	1	1	0	1	1	1	O Batuque Carioca, p. 31, quad-bell 2
1	1	0	1	0	1	0	1	1	1	0	1	0	1	0	1	3	O Batuque Carioca, p. 31, quad-bell 4
0	1	0	1	1	1	1	1	1	0	1	0	0	1	0	1	1	O Batuque Carioca, p. 33, cuica (levada 5)
0	1	0	1	1	1	1	1	1	0	0	1	1	1	0	1	1	O Batuque Carioca, p. 33, cuica (levada 6)
0	1	0	1	1	0	1	0	1	0	1	0	1	1	1	1	1	O Batuque Carioca, p. 33, cuica (levada 7)
0	1	0	1	1	0	1	0	1	1	0	1	1	1	1	1	1	O Batuque Carioca, p. 33, cuica (levada 8)
1	0	0	1	0	1	0	0	1	0	0	1	0	1	0	0	3	O Batuque Carioca, p. 55, G.R.E.S. Portela (caixa)
1	0	0	1	1	0	0	1	1	0	1	0	1	0	0	1	1	O Batuque Carioca, p. 60, G. R. E. S. Unidos de Vila Isabel (tarol)
1	0	1	0	1	0	0	1	0	1	0	1	0	1	1	0	2	Pandeiro Brasileiro, Vol. 2, p. 63, Samba/Choro 17
1	0	0	0	1	0	1	0	1	0	0	0	0	1	0	1	1	Ritmos do Brasil para Bateria, [no page numbers], Partido Alto 3
1	1	0	0	1	0	1	0	1	0	0	1	0	1	0	1	1	Ritmos do Brasil para Bateria, [no page numbers], Partido Alto 5
1	0	1	0	0	1	0	1	0	1	0	1	0	0	1	0	2	Ritmos do Brasil para Bateria, [no page numbers], Samba (cont.) 1
1	0	1	0	0	1	0	1	0	1	0	1	0	1	1	0	2	Ritmos do Brasil para Bateria, [no page numbers], Samba de Roda 9
1	0	0	1	0	0	1	0	1	0	1	0	0	1	1	1	1	Ritmos do Brasil para Bateria, [no page numbers], Bossa Nova 2
0	1	0	1	0	0	0	1	0	0	0	1	0	1	0	1	1	New Ways Of Brazilian Drumming, p. 30, "main rhythmic cell" of Partido Alto

Figure 70: Traditional samba patterns extracted from instructional books for the training set.

¹⁶⁴ The Spanish-language *Candombe* of Uruguay (with accent on the second syllable and no 'l') is somewhat different from the Portuguese-language *Candomblé* (with accent on the third syllable), but both have their roots in the music and religion of the *Yoruba* of West Africa.

Preliminary neural-net runs were conducted, and, as also found later on in other experiments, the neural nets were able to learn certain clave categories (the primary ones) quite successfully, but had difficulties with the neutral and incoherent categories. Specifically, among the 3152 vectors reclassified after the golden-ear trials, there were 558 members of category 0 (incoherent), 1313 members of category 1 (3-2 clave direction), 1101 members of category 2 (2-3 clave direction), and 180 members of category 3 (neutral) under the strict classification. In the best neural-net run, these outputs (with one-up output encoding) had learning rates of 0%, 90%, 89%, and 31%, respectively. When weighted by the proportions of the output categories, this gives a learning rate of 70% (the highest attainable generalization rate in Music Information Research to date¹⁶⁵), but not a great result for CI purposes.

In response, data files were created that had neutral patterns repeated such that their number was comparable with the numbers of 3-2 and 2-3 patterns. In addition, 82 more neutral patterns were identified and added to the training set. At the same time, the incoherent-category output was removed so as to test whether it was possible to learn the 3-2 and 2-3 directions without reference to the neutral or incoherent options (as humans are typically taught clave). In this output encoding, when necessary, the incoherent category is expressed by low output levels for both 3-2 and 2-3. With no immediate improvement in learning performance, experiments continued. (Note that these experiments are concerned only with learning, not generalizing. The purpose is to set the benchmark against which generalization performance will be tested at all subsequent stages of the research.)

In the fall of 2008, four sets of data were obtained, as shown in Table 17, where the neutral-category vectors were repeated in the sets marked ‘boosted’:

Also, at this point, it was necessary to update the output encoding, which was a hold-over from the early stages of the research before the incoherent classification was not included.

A better coding scheme that is consistent with two-bit binary coding would be *neutral–reverse–forward–incoherent*, based on Table 18.

Also under consideration at this point was a pseudo-one-up code featuring the neutral, forward, and reverse categories—those most commonly encountered in

¹⁶⁵ According to multiple private communications during the main conference of the Music Information Retrieval community, ISMIR, 2008, Philadelphia, PA.

learning about and playing clave-based music: 1–0–0 for neutral, 0–1–0 for reverse, and 0–0–1 for forward.

Looking for more ways to boost the training set by making use of musical knowledge and intuition, it was next decided to add mirror images (180° phase shift) and single-bit-shifted versions (22.5° phase shift) of the existing patterns to the classified data set (Figure 71). For patterns that are indubitably 3-2 or 2-3, it is known (as part of common and basic knowledge about the clave tradition) that a 180° phase shift reverses the clave direction. It is also known that a 180° phase shift would preserve the neutrality of fully neutral patterns. The effect of this on weak and incoherent patterns, as well as loosely neutral patterns was unknown. Similarly, it was discovered by the researcher early in his clave-teaching experience that a single bit shift (a 22.5° phase shift) would reverse the clave direction of most patterns in 2-3 and 3-2 direction¹⁶⁶, but *may or may not* switch a neutral pattern to incoherent and vice versa. Classification of these new patterns, preceded by a redundancy check (shifted patterns could already have been included in the data set as part of the random selection), was performed over another several weeks.

1 1 1 1	1 0 1 0	0 0 1 0	0 0 1 1	1 0 0 0	1	64035	ave.	1 1 1 1	1 0 1 0	0 0 1 0	0 0 1 1	1 0 0 0	1	64035	ave.
0 1 0 1	1 1 1 0	1 0 0 0	0 0 0 1	1 0 1 0	1	24193	very strong	0 1 0 1	1 1 1 0	1 0 0 0	0 0 0 1	1 0 1 0	1	24193	very strong
1 0 0 0	1 1 0 0	1 1 0 1	1 1 0 1	1 2 1 0	1	36061	ave.	1 0 0 0	1 1 0 0	1 1 0 1	1 1 0 1	1 2 1 0	1	36061	ave.
1 0 0 0	0 0 0 0	1 0 1 1	1 0 1 1	1 0 1 0	1	32949	strong	1 0 0 0	0 0 0 0	1 0 1 1	1 0 1 1	1 0 1 0	1	32949	strong
1 0 1 0	0 0 1 1	0 1 0 0	1 1 1 0	2 1 0 0	1	41806	very strong	1 0 1 0	0 0 1 1	0 1 0 0	1 1 1 0	2 1 0 0	1	41806	very strong
1 1 0 0	1 0 0 1	1 0 1 0	1 0 0 0	1 0 1 0	1	51624	weak	1 1 0 0	1 0 0 1	1 0 1 0	1 0 0 0	1 0 1 0	1	51624	weak
1 1 1 0	0 1 0 0	1 0 1 0	1 0 1 0	2 1 0 0	1	58410	strong	1 1 1 0	0 1 0 0	1 0 1 0	1 0 1 0	2 1 0 0	1	58410	strong
0 0 0 0	0 1 0 1	1 1 0 0	1 1 1 0	2 1 0 0	1	1486	strong	0 0 0 0	0 1 0 1	1 1 0 0	1 1 1 0	2 1 0 0	1	1486	strong
0 0 1 0	0 1 0 1	0 0 0 1	1 0 0 1	2 1 0 0	1	9497	very strong	0 0 1 0	0 1 0 1	0 0 0 1	1 0 0 1	2 1 0 0	1	9497	very strong
1 1 0 0	0 0 1 0	1 0 1 0	1 0 0 1	1 0 1 0	1	49833	strong	1 1 0 0	0 0 1 0	1 0 1 0	1 0 0 1	1 0 1 0	1	49833	strong
0 0 1 1	1 0 1 0	1 1 1 1	1 1 0 0	1 0 1 0	1	15100	ave.	0 0 1 1	1 0 1 0	1 1 1 1	1 1 0 0	1 0 1 0	1	15100	ave.
0 1 1 0	0 1 0 1	1 0 0 1	1 0 1 0	2 1 0 0	1	26010	very strong	0 1 1 0	0 1 0 1	1 0 0 1	1 0 1 0	2 1 0 0	1	26010	very strong
1 1 1 1	0 1 0 1	1 1 0 0	1 0 1 1	2 1 0 0	1	62931	strong	1 1 1 1	0 1 0 1	1 1 0 0	1 0 1 1	2 1 0 0	1	62931	strong
0 0 1 1	0 1 0 0	1 1 0 1	1 0 0 1	2 1 0 0	1	13522	strong	0 0 1 1	0 1 0 0	1 1 0 1	1 0 0 1	2 1 0 0	1	13522	strong
1 1 1 1	0 1 1 0	1 1 0 0	1 1 0 0	1 0 1 0	1	63081	ave.	1 1 1 1	0 1 1 0	1 1 0 0	1 1 0 0	1 0 1 0	1	63081	ave.
0 1 1 1	1 1 0 0	1 0 1 1	1 1 0 0	2 1 0 0	1	31836	very weak	0 1 1 1	1 1 0 0	1 0 1 1	1 1 0 0	2 1 0 0	1	31836	very weak
1 1 1 0	0 1 0 1	1 0 0 0	0 1 0 0	0 0 0 0	0	58628	very weak on 3	1 1 1 0	0 1 0 1	1 0 0 0	0 1 0 0	0 0 0 0	0	58628	very weak on 3
1 1 1 0	1 1 0 1	0 1 1 0	1 0 0 1	2 1 0 0	1	60777	weak	1 1 1 0	1 1 0 1	0 1 1 0	1 0 0 1	2 1 0 0	1	60777	weak
0 1 1 1	1 0 0 0	0 0 1 1	1 0 0 1	1 0 1 0	1	30771	weak	0 1 1 1	1 0 0 0	0 0 1 1	1 0 0 1	1 0 1 0	1	30771	weak
1 0 0 1	1 0 0 0	0 0 0 0	1 0 0 0	1 0 1 0	1	38920	weak	1 0 0 1	1 0 0 0	0 0 0 0	1 0 0 0	1 0 1 0	1	38920	weak
1 1 0 0	1 1 1 0	1 0 1 0	1 0 1 0	1 0 1 0	1	52901	strong	1 1 0 0	1 1 1 0	1 0 1 0	1 0 1 0	1 0 1 0	1	52901	strong
1 1 0 0	1 1 1 0	0 0 0 0	0 0 0 0	1 0 1 0	1	52736	very weak	1 1 0 0	1 1 1 0	0 0 0 0	0 0 0 0	1 0 1 0	1	52736	very weak
0 0 1 1	0 0 1 1	1 0 1 0	1 0 1 1	2 1 0 0	1	13163	very weak	0 0 1 1	0 0 1 1	1 0 1 0	1 0 1 1	2 1 0 0	1	13163	very weak
1 0 0 1	0 0 0 0	0 1 0 1	1 1 0 0	2 1 0 0	1	36956	weak	1 0 0 1	0 0 0 0	0 1 0 1	1 1 0 0	2 1 0 0	1	36956	weak
0 1 0 1	0 1 1 0	1 1 1 0	0 0 0 1	1 0 1 0	1	22241	strong	0 1 0 1	0 1 1 0	1 1 1 0	0 0 0 1	1 0 1 0	1	22241	strong
1 1 0 1	1 0 0 0	0 0 1 0	1 0 1 1	1 0 1 0	1	55339	strong (St.), very	1 1 0 1	1 0 0 0	0 0 1 0	1 0 1 1	1 0 1 0	1	55339	strong (St.), very
0 1 1 1	1 1 1 0	1 1 0 0	0 0 1 0	1 0 1 0	1	32354	very weak	0 1 1 1	1 1 1 0	1 1 0 0	0 0 1 0	1 0 1 0	1	32354	very weak
1 0 1 1	1 1 1 1	0 0 0 1	0 0 0 0	3 1 1 0	0	48912	weak on both!	1 0 1 1	1 1 1 1	0 0 0 1	0 0 0 0	3 1 1 0	0	48912	weak on both!
1 0 0 1	1 1 1 1	0 0 0 0	0 0 0 1	1 0 1 0	1	40705	weak	1 0 0 1	1 1 1 1	0 0 0 0	0 0 0 1	1 0 1 0	1	40705	weak
0 1 0 0	0 0 1 1	0 1 1 1	0 1 1 0	1 0 1 0	1	17270	strong	0 1 0 0	0 0 1 1	0 1 1 1	0 1 1 0	1 0 1 0	1	17270	strong
0 0 0 1	1 0 0 1	1 0 1 1	1 0 0 0	1 0 1 0	1	6580	strong	0 0 0 1	1 0 0 1	1 0 1 1	1 0 0 0	1 0 1 0	1	6580	strong

Figure 71: Discrete membership degrees in each of the clave categories arose during the classification of newly added patterns, shifted by 180° and 22.5°, according to the strict and loose interpretations.

¹⁶⁶ This conclusion, initially derived thanks to students who performed such shifts unintentionally, was independently (and without prompting) verified by an established and experienced Brazilian-music educator, J. Pegg.

Early into this re-classification session, a consistent qualitative commenting scheme arose in which the researcher would comment on the degree of each vector's membership in the given clave-direction category in terms of five membership degrees: very strong, strong, average, weak, and very weak. Multiple passes through randomly selected sections of the 6000 or so classified vectors verified that the comments were consistent from one pass to another.

Table 19: Breakdown of Classification Data by Category, Boosting, and Classification Mode

Data as of November 2008	Unique Set	Boosted Set
Strict Interpretation	558 in category 0; 1313 in category 1; 1101 in category 2; 262 in category 3.	558 in category 0; 1313 in category 1; 1101 in category 2; 1310 in category 3.
Loose Interpretation	191 in category 0; 1340 in category 1; 1102 in category 2; 601 in category 3.	191 in category 0; 1340 in category 1; 1102 in category 2; 1649 in category 3.
Total number of vectors	3234	4282

Following the completion of this last re-classification and data-set extension, a number of fully-connected multilayer perceptrons were created to train and test on the classified patterns to serve as a benchmark for evaluating the main research goal, which is investigating generalization-performance improvement under prestructuring. This is described in detail in the main body of the dissertation.

At this point, a significant concern may be raised about the scientific validity of the research if the author's expert knowledge of clave direction has gone through this many changes. Two answers to such a challenge are offered here:

The author's classification of those patterns that are either traditional or otherwise very clearly in one or the other clave direction did not change. Only the assessment of vague and non-traditional patterns which do not commonly arise in

included due to the statistical and technological nature of the research project, underwent any change.¹⁶⁷

This concern is only of consequence in the musical aspect of the research; not in the engineering aspect. As long as the “brain state” of the researcher-as-expert is frozen at a self-consistent mapping, and the artificial neural networks are trained to emulate that brain-state, and the engineering problem can be defined and solved.

Table 20: Justification for the new one-up output code based on two-bit binary coding for 3-2 and 2-3 clave direction.

2-3 OK?	3-2 OK?	overall category
no (0)	no (0)	incoherent
no (0)	yes (1)	forward (3-2)
yes (1)	no (0)	reverse (2-3)
yes (1)	yes (1)	neutral

The next few years saw continuing examination of the problem space, the nature of the *partido-alto* clave-direction problem, and the design of artificial neural networks.

The classification of the data was finalized (frozen at a reasonably reliable point) in July 2010 (year seven of the research), and the final experimental-design effort was begun.

E.3 Data Acquisition: Golden Ears

Some findings and conclusions from the golden-ear process follow:

1011 0010 1101 0100: incoherent¹⁶⁸;
1001 0101 1101 0111: 2-3;

¹⁶⁷ In connection with this justification of including atypical patterns in this research, consider the goal of musical grammars, as explained by Baroni: “the aim is to arrive at a set of rules that not only possesses coherence and concision, but also the necessary comprehensiveness, that is, one able to account for all phrases composed — or which one might want to compose — in the style of a particular repertory.” [7]

¹⁶⁸ Meaning there is no clave direction or no sense of clave in terms of the *partido-alto* interpretation of clave direction this research is concerned with.

0100	0100	1101	1111: incoherent;
0110	1100	1011	1101: 3-2;
1110	1100	1010	0011: loosely 2-3, strictly incoherent;
1011	1011	0101	0101: loosely 3-2, strictly incoherent (though in Afro-Cuban rumba, this would most likely be 2-3);
1111	0101	1110	0101: loosely 3-2, strictly incoherent;
1100	1000	1000	0001: 3-2;
1100	0001	0110	1011: 2-3;
0110	0101	1111	0011: 2-3;
0110	1111	1000	1101: 3-2;
1111	1001	0100	1111: 2-3;
1000	0101	1001	0010: 2-3;
0001	0110	0100	0001: a pattern that is so context-dependent, it has the potential to be in all four categories;
1001	0100	0011	0100: 3-2, demonstrating the relative nature of clave direction;
1111	0100	1001	0110: 2-3;
1000	0011	0010	1010: barely 2-3, loosely neutral;
1011	0111	0011	0110: 2-3;
1011	1110	1011	0001: 3-2;
0011	0001	1101	1101: 2-3, even though the last four onsets are typically associated with 3-2;
0001	0100	0010	1000: Strictly incoherent, but this pattern has a strong sense of clave (something I have called the “African sense”) because of the way the two cells of “101” (that are similar if viewed outside the grid of 16 th notes) are placed at 16 th -shifted positions with respect to one another, much like 1001 0000 0100 1000 (2-3) or 0101 0000 1010 0000 (3-2). Hence, the pattern follows the fundamental idea behind the generative root-notion of clave. However, this sense of displacement of identical cells at different reference points in the unraveling of musical time is here done in such a way that the Brazilian foundation of clave direction (the <i>partido-alto</i> and the “bossa clave”) are consistently violated. Nonetheless, in the loose interpretation, we would call this 3-2 because the resolution occurs on “and-four.”
0011	0110	1100	1010: clearly 2-3;
1110	0011	1010	1000: barely 3-2.

E.4 Bootstrap Process: Misclassification and NN Feedback

Preliminary neural-network runs gave rise to the question of whether there is a consistent correspondence between output levels and membership degrees (very strong, strong, average, weak, and very weak) of the input vectors. Thus the research questions governing the benchmark stage multiplied:

1. Which output encoding is the best for fully connected MLPs to learn clave direction? (This question is necessary from the point of view of scientific honesty: The output of the present research must be compared with the best that standard neural nets can do.)
2. What are the optimal (or near-optimal) numbers of hidden layers and hidden-layer elements for fully connected MLPs for learning clave direction?
3. Do the unquantized, non-thresholded (raw) network output values correspond with the degree of membership of patterns in the clave-direction categories as implied by the membership qualifiers very strong, strong, average, weak, and very weak? (And should they? After all, the networks are not presented with this information. Is it reasonable to expect the MLPs to infer this information from a randomized presentation of a mixed group of vectors, or should very strong vectors be presented earlier or more frequently than, say, weak vectors, and so forth?)

Multiple experiment-organization schemes were devised and carried out in order to answer these questions. The available vectors, after the most recent classification, were separated into not only strict and loose sets, but also into those that are very strong members of the various categories, as well as STRONG (defined above as strong and very strong), average-and-up members, and various groupings of weak members, as well as one other grouping: those vectors for which the strict and loose classifications are in agreement (“agreed”) versus those for which they are not.

All three output encodings were trained and tested with agreed STRONG vectors, agreed AVERAGE + STRONG vectors, strict AVERAGE + STRONG vectors, loose AVERAGE + STRONG vectors, all but the neutrals, all strict vectors, and all loose vectors. This was carried out within a multi-tier training scheme (three tiers in half the experiments, and four in the other half), first training only with STRONG examples (including strong incoherents, but not including any neutrals) for which the strict and loose classifications were in agreement. The resulting network was then trained by adding the average 2-3 and 3-2 patterns to the pre-existing training set, after

which, the source of the additional training vectors for different runs was split: AVERAGE + STRONG vectors for which the strict and loose classifications did not agree were added to the corresponding training set for a third round of training, which still included the initial training set, making the changes brought about by subsequent training passes smaller than if the networks had been reset and trained from scratch with the new sets. A final and fourth training pass was carried out on half the runs (to examine the difference between training only with strong examples, and eventually including all examples), adding all non-neutral patterns to the training set. This scheme was intended to emphasize the effect of the very strong and strong patterns while giving the weaker examples a small voice in the overall representation achieved by the networks. Since the output-encoding question had not been resolved, these runs were carried out for each of the three output schemes.

For the agreed strong and up, the binary, pseudo-one-up, and true one-up networks' learning rates were 92–98, 0–100–99, and 0–98–96–83, respectively. These and other results are tabulated below in Table 20. Binary encoding has two output “wires,” pseudo-one-up has three, and one-up encoding has four.

Table 21: Learning Rates per Output for Multi-Tier Training (multiple membership-degree groups in training and test)

	strict classification				lenient classification			
	good examples		all examples except neutral		good examples		all examples except neutral	
	Train (3 tiers)	Test (4 tiers)	Train (3 tiers)	Test (4 tiers)	Train (3 tiers)	Test (4 tiers)	Train (3 tiers)	Test (4 tiers)
binary encoding	94–97	81–93	80–96	77–96	94–97	87–90	92–95	84–95
pseudo-one-up encoding	0–99–98	0–95–95	0–96–95	0–96–95	0–99–98	0–94–95	0–95–95	0–95–95
one-up encoding	0–97–96–45	0–91–89–32	0–79–87–56	0–79–87–56	0–97–97–35	0–88–89–35	0–95–96–0	0–95–96–0

In order to choose the best combination, a simple Pareto-optimization¹⁶⁹ scheme was generated. The two criteria chosen were average percentage learning rate and the percentage of the least successful output unit. Since the final goal of the research is generalization performance on a test set, only the test results were considered. The training percentages were ignored. It is clear from both the average percentages and the least successful output-unit activation that the binary encoding scheme gave the best performance. (This is in direct opposition to what was stated above, and will be discussed subsequently.) Furthermore, the three-tier approach to training consistently gave equivalent or better results than when weak examples were included in training at the fourth step. However, in hindsight, this could also be due to the fact that the corresponding test set included those weak examples, which have continued to present difficulties in terms of classification. In fact, the refinement of the classification of weak examples continued past this stage of the research. Hence the suggestion that training on clearer, stronger examples leads to better generalization performance is not conclusive. Also it must be noted that due to the logistics of keeping track of this many neural-net runs, only the default absolute-winner method was used for determining these rates. A more reliable conclusion requires looking in detail at the output levels (in terms of thresholding as well as membership degrees) of each of these 24 networks. However, before that could be done, it was decided that a larger set of classified patterns was needed. (Approximately 3000 out of 65000 make about 5% of the I/O space, and even if randomly selected, better coverage is desirable.)

The new data set, as of mid-March 2009, after removing 542 duplicate lines, consisted of 8953 unique vectors classified and annotated. This version of the data made use of insights gathered during various rehearsals. It also made use of the lesson that spending all of one's time staring at binary vectors can lead to losing sight of the real-world problem behind the data set—the practical interpretation of clave-direction in samba performance—and that the path to acquiring this knowledge and being able to translate it into data is treacherous. For example, during the aforementioned rehearsals, the author realized that he had recently classified the vector 1000100010001000 as incoherent, thinking that it does not *establish* any sense of clave (which is true). However, what takes precedence in classification in terms of clave direction is the presence or absence of harmony with the given clave direction, and in that sense, it is a necessity of music-making that patterns such as 1000100010001000 be considered

¹⁶⁹ Pareto optimization is a decision-making method for multivariate problems. For example, automobile performance can be rated in terms of fuel-efficiency, safety, handling, cost, size, and other factors. In order to make a decision without discarding any of these factors, Pareto-optimality defines the concept of domination wherein a candidate solution dominates another if it is no worse than the latter in any respect, and better than the latter in at least one respect. Those candidates not dominated by any other candidate are said to form a Pareto-optimal frontier of solutions.

neutral to clave (and they generally are). In other words, ‘neutral’ is defined receptively, not generatively. This is an important distinction between the two troublesome categories (neutral and incoherent): The incoherent patterns actively go against the sense of clave established by *partido-alto* and the “bossa clave,” while neutral patterns passively accompany rhythms in clave without being in clave themselves. The incoherent patterns are recognizable through their three characteristics: (i) constant alternation of on-beat and offbeat nibbles¹⁷⁰, (ii) complete mismatch with the 1010 and 0101 sections of the *partido-alto*, and (iii) a very strong sense of rhythmic impropriety (in the narrow context of traditional *samba carioca* only, but still difficult to quantify).

In another rehearsal, a schema that had given the author much trouble during classification (1110) was recognized (when in the context of actual music) as more on-beat than not, thus helping draw any section it appears in toward the 2-side of clave.

Several other vectors whose classification could raise questions about consistency were identified in the course of reviewing the data, some based on the zipper analogy developed by the author while teaching samba during one of his regular weekly classes. The zipper analogy relates to the observation that when a pattern matches a standard Brazilian pattern (such as *partido-alto* or *teco-teleco*) for about a nibble’s duration, the pattern in question is highly likely¹⁷¹ to be in clave accordance with that standard pattern.

Two more passes over all of the data followed. The 542 duplicate patterns were checked for consistency, and some inconsistencies were found, primarily in the weak or incoherent patterns.

It is perhaps disappointing, but ultimately explicable, that there were some inconsistencies in both annotation and classification. Not having the vectors in randomized order (as they have to be when conducting experiments) allowed the researcher to consider neighboring patterns with respect to one another, and observe the effect of incremental changes (such as bit flips) that at times do not alter either the classification or the annotation, at times alter only the annotation, and occasionally change a vector from one category to another. (Much of this was understood at earlier stages in the classification process, as mentioned above when discussing single-bit shifts,

¹⁷⁰ Nibble: a string of four consecutive bits

¹⁷¹ This is highly likely but not certain because of the relative nature of clave: The entire pattern must be considered before a decision can be made. With all possible patterns, it is necessary that some patterns exist which match *teco-teleco* perfectly for about a nibble’s worth, but alternately rapidly between on-beatness and offbeatness for the rest of their course. An example follows: 1011 0101 1010 1011. The portion that matches the 3-2 *teco-teleco* zipper-style is shown in italics. This pattern is not in 3-2 even though the zipper analogy might suggest so. As with all clave-direction-discerning heuristics, the zipper is not a guarantee. All other criteria and the full length of the pattern must be considered.

but new insights into the same aspects of clave direction were acquired nonetheless.) Some of the inconsistencies observed led to reviewing many of the non-duplicated patterns as well, sometimes looking at entire blocks that shared one or two nibbles.

The most common inconsistencies were with patterns that contained the nibble 0110, which appears as a schema for the so-called 3-side, but depending on its neighbors, can bolster the sense of the opposite clave direction in many cases. It is thus the most context-dependent of schemata—it features the two most important onsets which each pull the schema in opposite clave directions. (In fact, in *all* MLP runs, the highest contributions to the output activation come from the *e* and *and* inputs, which are 0110|0110|0110|0110).

Also during this process, another implication of the strict/loose distinction became apparent through the examination of incoherent patterns: Members of the forward and reverse classes in the strict classification correspond to mainstream core traditional samba patterns—those typical patterns that can be found commonly in *samba carioca*, *samba do morro*, *samba batucada*, *pagode*, and some *Candomblé*—while the loose/lenient mapping encodes the relative nature of clave into a set-extension of traditional clave direction to include arbitrary non-traditional patterns that could conceivably arise in improvisation and perhaps through the influence of other musics.

As a result of all these clarifications, most (but not all) cases where patterns were previously classified as incoherent *due to being very sparse* were reclassified. Since the author has identified numerous pitfalls in the classification process (but mainly with regards to incoherent and neutral patterns, and weak examples), it was decided that other samba and clave experts would be consulted with regards to the category-0 and category-3 patterns, especially the ones that were annotated as weakly so, or were recently reclassified. Because the three international experts were not available during the winter of 2009, five local musicians with the greatest amount of experience with clave (many of them comparable to the author's, and one regarded by the author as the only local musician who surpasses his level of clave focus) were contacted. Only three were ultimately able to schedule time for these sessions.

Along with the two target experts, the same experimental setup was first executed with the three other valuable musicians, whose trials were to serve as preparatory for how to (and how not to) run trials with the two leading local experts. In the meantime, the processes of sorting out categories of vectors for different training and test tasks, reviewing the data for inconsistencies, and designing and running various MLPs and (new at this stage) Self-Organizing Maps (SOMs) continued.

On March 10, 2009, a new MLP input file was created that included very strong and strong category-1 vectors, very strong and strong category-2 vectors, and very strong category-0 vectors. Another training set was constructed with the addition of 50 fully neutral vectors (very strong examples of category-3).

According to the selection criteria in NeuralWare, the best network in each output-encoding category learned the training sets to the extent shown in Table 21.

Table 22: Learning comparison for the three output-encoding schemes (with percentages per output “wire”) and training with and without neutral vectors for strong examples.

TRAINING WITH STRONG EXAMPLES OF CATEGORIES 0, 1 & 2			TRAINING WITH STRONG EXAMPLES OF ALL FOUR CATEGORIES		
TWO-BIT BINARY	THREE- NODE	TRUE ONE-UP	TWO-BIT BINARY	THREE- NODE	TRUE ONE-UP
98–99	82–97–97	0–97–98–85	91–100	0–100–100	0–98–99–83

Also during this session, it was discovered that at some prior sorting stage, the annotations for the vectors became disconnected from the vectors, probably due to a fault in column selection when sorting. Another review of the data was triggered by this accident because the errors that would result from such a scrambling of inputs and outputs cannot be tolerated.

The foremost research question at this stage was identified to be the following. Can categories 0 and 3 be inferred from the output-activation levels of networks trained on categories 1 and 2 (as humans are typically trained in clave matters)?

In addition to the training and test data described above, a training set of only the very strong examples of categories 1 and 2, and two test sets, one consisting of only strong neutral patterns and one consisting of only strong incoherent patterns were generated. These sets were sorted not only by clave-direction, but by membership degrees as well.

The classification rate is the appropriate NeuralWare instrument for one-up encoding, and the confusion matrix is the appropriate instrument for other representations (such as binary encoding). Since different results were observed in past runs for choosing different output elements for the “save best” criteria, multiple “best nets” were selected, one with each appropriate criterion.

For the true binary output scheme, as if to make the extra effort seem pointless, either classification matrix gave the same best result: 99.99% and 99.91% learning on the two inputs for the file including only strong and very strong category-1 and category-2 vectors. Not only is the learning rate essentially 100%, but the low output activations are all 0.1 or lower, and the worst (lowest) high is 0.8.

Moving on to the three-element output (the pseudo-one-up code), and using confusion matrix 3 as the criterion, there is no degradation in results from the binary case, except that category-3 is learned to 0% (as expected). However, the highest low output level is 0.19 and the lowest high output level is 0.78.

There was one category-0 (output 000) vector that made it into this data set by mistake, and it came out 001. The vector is 0101|0101|0110|1101, which should be category-0 for the strict and strong category-1 for the lenient classification, and it was classified as a strong category-1 here, which given the remaining 2599 strong examples, and the fact that no distinction has yet been made to the network about strict and loose classification, is a good result. This was vector 21869. Looking through the data set, vectors 21793, 21816, 21833, 21865, and 21868 were also fixed for inclusion in the wrong data set.

For the true one-encoding, the classification results were again the same, and the only change in output levels was that the lowest high was 0.8076.

At this stage, no preference for any one output-encoding scheme has been identified, and by nature of the data used up to this point, membership degrees have not been explored. However, there are some indications that the membership degrees are reflected in the output-activation levels; consider the following strictly incoherent patterns: 0010|1011|1111|1100, which came out with 0.80 category 2 and 0.19 category 1, and 0111|1011|1101|0110, which came out with 0.55 category 2 and 0.43 category 1. Although neither pattern is in the 2-3 direction, the former is closer to being in 2-3. However, looking at some of the neutral patterns and how they were classified by the current network, it is highly puzzling that 0000|0000|0000|0000 should come out with output activations of 0.67 for category 2 and 0.33 for category 1. Since the input values are all 0, only the bias can be responsible for this, which suggests that the bias connection is moving the network toward 2-3 clave direction when there is no other basis for a decision. On the other hand, the neutral pattern 0000|1101|0000|1101 came out with output activations of 0.30 for category 2 and 0.67 for category 1. If the previous conclusion is true, that implies that inhibitory connections (negative weights) may be responsible for this change. Besides, the nonlinear nature of multilayer perceptrons renders further conjecture meaningless. The important thing is to employ good, theoretically supported design practices to achieve the best-performing network(s). Furthermore, it is possible that anecdotal examples of both success and failure can be identified for the membership-degree question; what is more important is to show consistent evidence in this regard. Once again, much more detailed analysis of *all* neural-net outputs is needed. These preliminary experiments mainly serve to identify questions, problems, and the best route to take in the final design of the fully connected network that will serve as the benchmark for the primary research. Another such question raised by the above examples (as well as the pattern 0000|1111|0000|1111, which has activations of 0.35 and 0.62), is whether some preference for the start of the 16-bit patterns has somehow been encoded, rather than the implicit understanding that these patterns are “infinitely looped.” Since the 180°-shifted version of *every* pattern in the original training set was included in the current

one, the “infinite looping” of patterns should effectively be realized by the shifting, *unless* the subsequent addition of single-bit-shifted patterns broke the symmetry to a sufficiently drastic extent.

At the next stage, strong examples of category-0 (“good zeros”) were added to the training set. The “best net” came out to be one that had all the vectors desired to be classified as category 0 instead classified as category 1. However, during training, many intermediate networks were observed when the desired/actual category-0 percentage hovered around 80, with the accompanying percentages for categories 1 and 2 not very low. Even after setting the “best net” criterion to confusion matrix 4, the software clearly did not do so. Why these networks were not selected by NeuralWare is among the technical (not scientific) problems to be solved during subsequent research. The use of the “check points” feature was then taken under consideration.

At the following step, the networks were trained on categories 1 and 2 only, but tested on categories 0 and 3 only (together for the former, separately for the latter). On the best one-up network trained only on strong and very strong category-1 and category-2 vectors, the category-0 test set was classified 52% into category 1 and 48% into category 2. Furthermore, the output activations were all over the map, with the majority very close to 0 or 1. From this we can conclude that simply training the network with the best examples and testing with the vaguest and in-between (anti-clave) examples does not empower the network to induce the fundamental difference between clave patterns and non-clave patterns. Some other method must be devised for the networks to learn to make “fuzzy” decisions.

Testing on fully-neutrals only, with the binary output encoding (when trained on strong and very strong category-1 and category-2 examples), the best network according to confusion matrix 1 classified 32% of the neutrals into category 1 and 68% into category 2. The best network according to confusion matrix 2 classified 38% into category 1 and 62% into category 2. Examination of the output activations of the latter suggested that the trained network had fully induced the symmetry in clave (symmetry to the extent that a 180° phase shift leads to the reversal of clave direction). However, this clearly did not help the networks learn the concepts of clave-neutrality and of a lack of clave sense. It appears that more than just the best examples of category 1 and category 2 are needed in order for the networks to learn more subtle aspects of clave direction. In fact, the networks that show this output pattern may have learned the clave notion so thoroughly that they clearly have difficulty imposing the human (psychomusical) concept of clave-neutrality (discussed explicitly in [4, p. 80] and implicitly in [4, p. 74] and [5, p. 10] on the well-defined concept of discrete clave directions.

Another puzzling, yet potentially useful observation is that the category-3 vectors were more likely to give output activations away from the upper and lower

extremes of the output-activation range (“rails”), whereas the category-0 patterns tested on the networks trained the same way were almost all clamped to the rails. It is reasonable to expect neutral patterns along the lines of 1000|1000|1000|1000 to fall near the halfway mark in both output activations, whereas patterns like 0110|0110|0110|0110 would be expected to excite outputs equally strongly, thus leading to plots more like those observed for the incoherent vectors. Aside from the need to carefully analyze the above results for differences between the vectors that had outputs around 0.5/0.5 and those that had outputs around 0/1 or 1/0, it is also necessary to devise a training method that will build the idea of the neutral category into the networks’ mappings. This raises the further question (once again) of the difference between strict and lenient classification: a pattern like 1010|1010|1010|1010 is classified as incoherent in the strict interpretation, but neutral when lenient.

Looking closely at the patterns whose outputs were close together, and those whose outputs were far apart (closer to 0 or 1) reveals no rhyme or reason for this behavior. The following patterns were found with output activations far apart:

```

0010|0000|0010|0000
0000|1011|0000|1011
1100|1001|1100|1001
0010|0011|0010|0011
1011|0110|1110|1110
1010|1111|1010|1111
0010|1111|0010|1111
0010|0010|1100|1100
1001|0010|1001|0010
1110|0010|1110|0010
1010|0010|1010|0010
1011|1000|1011|1000
0001|1001|0001|1001
0000|1001|0000|1001
1010|1010|0011|0011
1010|1000|1010|1000
1001|0100|1001|0100
1000|1110|1000|1110
1011|0110|1010|0011
1111|1111|0010|0010
0111|0111|1101|1101.

```

On the other hand, the patterns whose output activations were clustered near the midpoint were:

0111 | 1111 | 0111 | 1111
 1111 | 1111 | 1111 | 1111
 0000 | 1111 | 0000 | 1111
 0011 | 0011 | 1110 | 1110
 0000 | 0000 | 0000 | 0000
 1111 | 0001 | 1111 | 0001
 1111 | 1001 | 0001 | 0001
 0111 | 0111 | 1010 | 1010
 1111 | 1011 | 1111 | 1011
 0000 | 1101 | 0000 | 1101
 1000 | 1111 | 1000 | 1111.

As to why 0111 | 0111 | 1101 | 1101 and 0111 | 1111 | 0111 | 1111 should belong to different output-activation-behavior categories (as classified by the NN), the author has no logical answer.

Examining the three-node versions for the same training and test sets (with only confusion matrix 3 as selection criterion since it is not clear whether this has any appreciable effect), it is found that category 1 and category 2 are again learned to a most excellent degree by the network trained on those categories.

Testing the same network on strong incoherent vectors and strong neutral vectors, the same results as before were obtained, showing inconclusive output activations for strong incoherent vectors on a three-node-output network trained on strong and very strong clue examples.

Nonetheless, some of these output activations indicate clues that the fuzzy interpretation may be workable. However, far more detailed analysis is needed with a much bigger sample of vectors and their outputs for this to be reliably concluded.

When the true one-up output encoding was used with the same data (strong and very strong examples of the two main clue categories), and tested on the “good zeros,” an interesting result was obtained:

The results in Figure 72 are not as expected. No pattern was classified as belonging to category 0. This is taken to be an indication of the highly nonlinear nature of clue direction: Neutral patterns can accompany both 2-3 and 3-2 patterns, but *not because they are made up of equal elements of those clue categories*. In fact, neutral patterns can be made up of all category-2-type schemata, all category-1-type schemata, a mixture of the two, or even neither. Furthermore, some patterns made up of all category-1-type schemata (and all category-2-type schemata) will necessarily be incoherent (category 0) in the strict classification and possible neutral (category 3) in the loose classification. Hence, even if the network(s) learn combinations of onset patterns that hint strongly at one clue direction or the other, because of the way clue direction works, they wouldn’t necessarily be able to discriminate neutral from incoherent on the basis of that

information alone. The incoherent and neutral categories must be part of the training set, and in comparable numbers to category 1 and category 2. The balance of strong and weak examples must be cared for. The question of how to deal with context dependencies that fall in between the extremes of the strict and loose classification schemes—which relates to a potential fifth category: don't-know, or highly context-dependent—needed to be resolved. Another question about membership degrees was still open: Can the networks learn membership degrees from being trained with only strong examples and being presented weak examples, or should such membership degrees be worked into the vector representation?

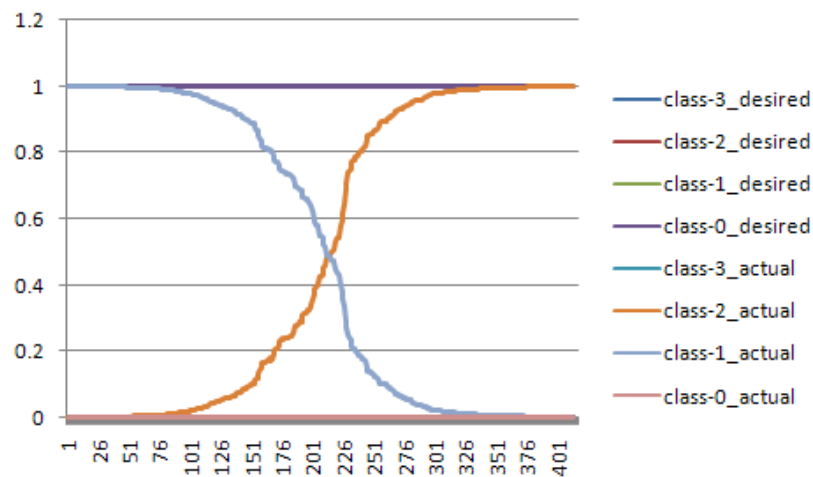


Figure 72: One-up (output encoding) network trained on good examples of category 1 and 2, tested on good examples of category 0 reveals a tendency to distribute the outputs evenly between categories 1 and 2, with a smooth transition.

These and other questions depend on resolving two issues: (i) Is the author's understanding of weak and incoherent patterns consistent? Is it possible for it to be consistent given the highly abstract representation and the context-dependent nature of the patterns, or should a category of high context-dependence be admitted? (ii) How can the necessary NN experiments be organized so that the number of runs and the analysis of their data are not insurmountable?

The second question has to do with the many experimental parameters: three (or more) output-encoding schemes, up to four "best net" criteria per output-encoding scheme, different numbers and sizes of hidden layers to be tried out, multiple training/test schemes that involve training each of the above combinations on different subsets of the data and testing them on different subsets of the data. Based on the experiments so far, the number of combinations is approximately 1200, each with

thousands of vectors whose output levels need to be scrutinized. Moreover, the musicological question, *i*, has to be resolved first in case all these efforts are for naught. The resolution of that question can increase or decrease the number of NN trials needed. Once that question is answered (*if* it can be answered), the experimental issues (mainly output-encoding selection and membership degrees) need to be prioritized. Perhaps, choices have to be made (heuristically, or even arbitrarily) on some of the other experimentation parameters.

At this point, the experimentation process was re-organized and re-prioritized. The resulting plan appears below in priority order. It is important to note that all these levels of reinterpretation and review are not “fudging” the data, but nailing down a precise understanding of the carioca clave-direction notion the author developed through experiencing standard traditional carioca samba, and has had to extend to uncharted realms of rhythmic possibilities. At each reclassification stage, clearer, more precise ideas of what constitutes a weak example of a certain clave direction, and what constitutes an example of neutrality or lack of clave direction are revealed.

1. Review the information on NeuralWare NeuralWorks instruments for selecting the "best net" criterion.
2. Analyze the two sets of pre-golden-ear results.
3. Review the existing classification and annotation for weak vectors, 0s and 3s based on any new findings from golden-ears or pre-golden-ear trials, as available.
4. Learn the NeuralWorks scripting language to automate future neural-net experiments.
5. Check the 50x50 SOM results to see if the triple separation corresponds with the result of step 2 above, or the idea of a fifth category.
6. Compare hidden-layer sizes in existing trials to choose hidden-layer size.
7. make sure that the 180-degree-reflected versions of the 22.5-degree-shifted patterns are in the data set.
8. Repeat category-0 and category-3 vectors in the data set until all categories are presented in comparable numbers.
9. Make a *modular* neural network for identifying a possible performance ceiling for the fully connected MLP. This network will have separate sets of hidden-layer elements with dedicated ‘yes’ and ‘no’ (one-up) outputs for every clave category. Full 16-bit input-pattern connections will be supplied to each hidden-layer group, but the groups will have no cross-connections. Thus, the membership of a pattern in one category, such as category-1, will not affect the network’s training for the same pattern’s membership in category-0. The trade-off is likely to be that the total symmetry will be compromised, but this is acceptable if the psychomusical concept of clave-neutrality can be encoded.

10. Reduce the training/test schemes to a reasonable number that makes sense from a clave-teaching point of view.
11. Whenever you analyze data, consider *every important question and significant aspect*. Look at every single misclassified entry; observe all output levels; threshold and re-plot. Do fewer trials, but *thoroughly analyze all results*.

Analysis of the two pre-golden-ear trials indicates that the results are not admissible. This conclusion is based on the finding that the musicians who took part in the pre-trials failed to identify clave direction correctly in *all eight* of the control patterns. These eight control patterns were chosen for being very clearly in one clave direction or another. One subject identified five out of eight correctly; the other six out of eight. Since these patterns were very basic, a subject needs to classify all eight for other responses to be admissible. These subjects—two of the five remaining best samba players in Portland—appeared to lack a highly critical and holistic understanding of the carioca-music interpretation of the clave concept. However, valuable insights to the process were gained, should the two true experts become available in the future for golden-ear trials.

With all new trial data inadmissible and other expert samba musicians unavailable due to being out of town, and with deadlines looming, the author turned once again to the original golden-ear trials with the three international experts, along with all of the notes, observations, and insights gathered over the past 7.5 years of this research.

All reviews of vectors annotated as average, strong or very strong category-1 and category-2 performed over the recent months have shown that these classifications and annotations are consistent; there is no question about them based on musical context. The remaining and relatively few patterns which fall into the categories of weak and very weak examples, and some of the cases of neutral and incoherent vectors, however, still need classification. This need not cause unease about the scientific validity of the research presented because no machine intelligence may be expected to mimic human understanding of a subtle cultural concept until a sufficiently simplified and abstracted version of that concept can be consistently identified by the human brain which the machine intelligence is expected to learn from. In this case, the process of classification and annotation has revealed to the author that clave direction is considerably more complex than even he thought at the beginning of this research, and further refinements are necessary if a consistent, repeatable, and reliable representation of clave direction is to be attained. This representation is the classification data set. The several steps of refinement already experienced were a necessary part of the learning process of performing a set-extension of the traditional samba notions of clave direction found in traditional patterns to all arbitrarily selectable patterns. One more

refinement is now performed, necessitated by the reflection of the author's changing ideas of the classification of very vague patterns in the results of the early neural-net trials. If the author's classification of highly vague patterns were not found to be context-dependent, then the failure of the networks to learn certain aspects of clave direction would have to be noted and accepted as the necessary ceiling to NN performance in this field.

However, since the author has identified (a small number of) cases wherein the multi-criteria approach used for classifying the data led to different answers at different passes for the most susceptible (most vague, most context-sensitive) patterns, it is only fair to the machine intelligence sought here to make this classification process more consistent through further narrowing down of the musical-contextual scope of the current definition of clave direction.

This was accomplished through a repeated analysis of the original golden-ears' comments and a full overview of all the thoughts developed and insights gained over the entire research process. The results have to do with the strict and loose/lenient classification schemes as well as the classification of vague patterns.

E.5 Bootstrap Process: Context-Dependent Patterns

The strict and loose classification schemes encoded several ideas at once, which is thought to be the reason for the difficulties in nailing down the vague patterns' classifications. 'Strict' implied that patterns were acceptable in the given clave direction in any musical context such as tempi, genres, and instrument choices. At the same time, they are repeatable (loopable). Furthermore, patterns that are strictly 2-3 are those that have no severe mismatch to *partido alto*. A severe mismatch is defined here as a 1010 schema coinciding with a 0101, or vice versa. 'Loose' implied that a pattern may not be 2-3 or 3-2 in every context, but it would be one context or the other, *or* that it would be acceptable in a certain clave direction is passing, i.e. not repeated or looped.

E.6 Final Approach

Eventually, the pursuit of looking for a fuzzy-membership interpretation of the NN output activations was abandoned in favor of the central research goal. Nevertheless, this process has led to better understanding of the data, the operation of the networks, and of clave direction itself.

References for Appendix E

- [1] IEEE Spectrum, September 2009, p. 10.
- [2] Lehmann, B. *The Syntax of 'Clave' – Perception and Analysis of Meter in Cuban and African Music*, Tufts University (Master's thesis), Arlington, Massachusetts, 2002.
- [3] Nzewi, M., *African Music: Theoretical Content and Creative Continuum*. Oldershausen: Institut für Didaktik Populärer Musik., 1997, p. 24.
- [4] Cruz, T., Moore, K., Gerald, M., and Fiol, O., *The Tomás Cruz Conga Method, Volume II—Intermediate—Essential Cuban Conga Rhythms*, Pacific, MO: Mel Bay Publications Inc, 2004.
- [5] Mauleón-Santana, R., *101 Montunos*, Petaluma, CA: Sher Music, 2005.
- [6] Gonçalves, G., and Costa, O., *Aprendendo A Tocar O Batuque Carioca – As Baterias das Escolas de Samba do Rio De Janeiro*, Brazil: Groove, 2000.
- [7] Baroni, M., “The Concept of Musical Grammar,” *Music Analysis*, Vol. 2, No. 2, 1983, pp. 175–208.

APPENDIX F: Brazilian Glossary, Clave Terminology, Musical Definitions, and Samba Concepts

F.1 Author's Original Samba Glossary

This glossary gives the author's own descriptions of samba instruments, ensembles, and elements of the surrounding culture and context. It is based completely on the author's experience of learning samba from Brazilian and other masters and teachers, as well as interacting with Brazilians and Brazilian culture in the US. Only one entry comes directly from another source, and that is marked.

A

Ala (*ah-lah*): 1. Wing; 2. Section of percussionists who play the same instrument in a *bateria*. Example: "ala dos tamborins" (tamborim section).

Alfaia (*ow-fah-yah*): Wooden-shelled, rope-tuned bass drum used in *maracatú*.

Agogô (*ah-go-go*): A dual-bell of African origins, consisting of a smaller, high-pitched bell, a larger, low-pitched bell, and a somewhat flexible connecting metal handle, which allows the two bells to be "clicked" together for grace notes; played with a drum stick on one hand and by squeezing (clicking) with the other hand. Sometimes, an *agogô* can consist of three bells or more, though more than three are not practical unless they are arranged in the form of the Império Serrano quad-bell. (See below.)

Angola (*ahn-go-lah*): 1. A country on the southern west coast of Africa, and the homeland of many, though not all, African slaves imported into Brazil by European slave-traders. 2. A form of slow *capoeira*, played close to the ground, and mostly developed prior to the acceptance of the martial art by the Brazilian authorities and its subsequent popularity as a tourist attraction and cultural export of Brazil. 3. One of two main branches of the Afro-Brazilian *Candomblé* religion.

Apito (*ah-pee-too*): The tri-tone samba whistle used for intra-*bateria* communication and embellishment.

Atabaque (*ah-tah-bah-key*): A family of rope-tuned ceremonial or sacred hand drums originating in West Africa and the transatlantic slave trade, the *atabaque* are three in number. They are similar to congas, but their sizes vary by height instead of girth. *Atabaque* are played with one hand and one stick (*Angola*

tradition of *Candomblé*) or two sticks (*Ketu* tradition of *Candomblé*). The three atabaques are called *rum*, *rumpri*, and *lê*.

Axé (*ah-shay*): 1. A Yoruba word for life energy, adapted to Portuguese spelling. 2. Energetic, fast, joyous, loud popular music style from the northeastern states of Brazil, especially Bahia. 3. The pop-music form of samba-reggae, as exemplified by Daniela Mercury, Luiz Caldas, Yvete Sangalo, Terra Samba, Cheiro De Amor, Gang do Samba, Bom Balanço, Banda Eva and others.

B

Baião (*bah-yown*): Northeast Brazilian cowboy music, played with a trio of triangle, bass drum and accordion. A very common rhythm used in Brazilian jazz, the baião is usually mistaken for samba.

Baixo (*bi-shu*): 1. Low; down 2. Bass 3. Bass guitar (electric bass). When a *surdo* is not present, the bass plays the parts of the first and second *surdo* (sometimes third as well) while following the chord changes.

Balança (*bah-lahn-sah*): 1. *Balança* means *swing*, as in a playground. The contour of offbeatness in samba resembles the path followed by a swing: high at the ends, and low in the middle.

Bandeira (*bahn-day-rah*): Flag; *Porta Bandeira*: Flag carrier for the samba school.

Bandolim: (*bahn-doh-leen*): String instrument used in *pagode*.

Banjo (*bahn-zhoo*): String instrument used in *pagode*.

Bass-o Nova (*bah-so nova*): A style of electronic samba that evolved toward the end of the '90s from trip-hop and bossa nova, using samba samples.

Baqueta (*bah-kay-tah or vah-kay-ta*): 1. Drum stick. 2. Flexible tamborim stick for *batucada*.

Bateria (*bah-teh-ree-ya*): Ensemble of hundreds or thousands of drummers playing the styles of samba called *batucada*, *samba de enredo* or *samba do morro*.

Batucada (*bah-too-kah-dah*): Percussion samba; an essential aspect of the *carnaval* parade; also “drum jam.”

Batuque (*bah-too-key*): 1. Percussive Brazilian music. 2. A type of partially improvised dance drumming from Recife, played on *atabaques*, and analogous to the Cuban *rumba guaguanco*, where the highest-pitched drum (unlike in *Candomblé*) is the lead that improvises in conjunction with the dancer.

Batuqueiro (*bah-too-kay-roo*): Someone who plays *batucada*.

Berimbau (*be-rim-bau*): Musical bow and primary instrument of *capoeira*, consisting of a gourd resonator, a taut metal wire, a thin stick, a flat stone, and a shaker basket named *caxixi*, thought to be of African or Asian origin. The *berimbau* requires four simultaneous movements to play: The gourd (resonating cavity) is pulled into and away from the belly to control resonance and give the instrument its voice-like quality; the rock is pushed to mute the bow, which is played by a thin stick. The support hand also holds and plays the *caxixi* (basket).

Bloco (*blow-coo*): Larger than a *bateria* and smaller than a samba school (*escola*), a *bloco* consists of dancers and drummers. *Blocos* in Bahia, which mainly play *samba-reggae* are known as *blocos afros*.

Bombo (*bohm-bo*): Metal-shelled bass drum used in *forró*.

Bossa Nova (*boh-sah no-vah*): A soft samba derivative from the early part of the 20th Century, bossa nova became famous throughout the world in the '50s and underwent a transformation as it incorporated more and more elements from Jazz.

Brasileira (*bra-see-lay-rah*): Brazilian (female noun, feminine adjective).

Brasileiro (*bra-see-lay-roo*): Brazilian (male noun, masculine adjective).

C

Cabasa (*kah-bah-sah*): Also known as cabasa, but not to be confused with the LP afuche/cabasa, a hollowed-out, decorated gourd, played by shaking; also known as *xequerê*, which is pronounced the same as the more familiar Cuban *shekere*.

Cabeca (*kah-bay-sah*): Head. “Da cabeça”: from the top.

Caboclo (*kah-boke-loo*): A person of Native and African heritage.

Cachaca (*kah-shah-sah*): Sugar-cane liquor.

Caipirinha (*khy-pee-reen-yah*): A mixed alcoholic beverage made with plenty of crushed lime, sugar and sugar-cane liquor, *caipirinha* is known to make a *sambista* even out of gringos.

Caixa (*kah-sha*): 1. Box 2. Cashier 3. A type of Brazilian snare drum, constructed and played rather differently than a European snare.

Caixa de Guerra (*kah-sha gee gay-huh*): 1. Deep, waist-worn “battle-style” *caixa*.

Caixa em Sima (*kah-sha een see-ma*): *Caixa de guerra* played “up above,” by balancing the drum on one arm.

Caixa Malacacheta (*kah-sha mah-lah-kah-shay-ta*): Deep, 12” samba snare.

Caixa Tarol (*kah-sha tah-roll*): Shallow and wide samba snare; plural *tarois* (tah-roysh).

Candomblé (*kahn-dome-blay*): A Catholic-Yoruba syncretic religion with roots in the Yoruba people of modern-day Nigeria and Benin, *Candomblé* is one of half a dozen or more syncretic religious traditions that evolved among African slaves in the Caribbean and South America. The music of *Candomblé* (regardless of “nation”), is played with a single bell and a family of atabaques. Styles include the *capula* and *barravento*. It is closely related to the *Candombe* (kahn-DOME-bay) tradition of Uruguay and the *bata* of Cuban *Santería*.

Capoeira (*kah-poo-ay-ra*): A martial art, originally from Angola, *capoeira* was practiced secretly for hundreds of years by Brazilian (African) slaves, whether they came from Angola or not. In order to keep the nature of *capoeira* secret from the slave-owners, it was disguised as a dance. *capoeira*, as evolved in Brazil, is also the root of American *break dancing*.

Carioca (*kah-ree-oh-kah*): Not to be confused with *karaoke*, a *carioca* is a person from Rio De Janeiro. However, according to *carioca* author Castro, anyone who feels like a *carioca*, and lives like one, is a *carioca* [1].

Carnaval (*kah-nah-yau*): Literally meaning “Put aside the meat,” *carnaval* is the pre-Lent celebration lasting 4 days and featuring 24/7 music, parades, and constant partying.

Carnavalesco (*kah-nah-vah-less-coo*): The lead visual designer for a samba school.

Carreteiro (*kah-hay-tay-roo*): Also known as *repinicado*, this is the rolling-16th feel of samba, especially as played by the *ala dos tamborins*.

Cavaquinho (*kah-vah-keen-you*): A four-string guitar-shaped acoustic instrument with steel strings, slightly larger than an *ukulele*, with a different tuning than either of said instruments, the *cavaquinho* (a.k.a. *cavaco*) is essential to both *chorinho* and *samba de enredo*.

In *sambas de enredo*, the *cavaco* mostly plays *partido-alto*; in *chorinho* music, it plays elaborate melodic lines. One may conjecture (based on linguistic and historical patterns) that the *cavaco* may have come to Iberia through the Moors, and may be related to the Turkish string instrument *kabak kemane* meaning "gourd violin."

Caxixi (*kah-she-she*): A woven basket with a flat bottom, partially filled with seeds, the *caxixi* hangs from the stick hand of a *berimbau* player and provides part of the complex rhythm of the instrument.

Charuto (*sha-roo-to*) : An older style of repinique with a wider shell and natural head that is coming back in fashion in the *baterias* of Rio.

Chocalho (*shu-cow-you*) : A type of shaker with large, metal jingles on a metal frame, the *chocalho* is a surprisingly loud instrument that is essential for delineating verse-chorus-refrain changes in the *samba de enredo*.

Choro / Chorinho (*shoh-roo / shoh-reen-you*): A apparent synthesis of samba, Jazz and Chamber music, *chorinho* is said to have developed independent of, and earlier than, American dixieland jazz. Almost always instrumental, it manifests itself in many forms, ranging from string-ensemble and string/flute music to what sounds like dixieland played to a samba feel.

Comissão de Frente (*co-me-sau gee frenchie*): The "Front Commission," one of the main competitive aspects of a samba school; the first impression and introduction to the year's theme's theatrical element.

Coro (*koh-roo*): Choir, chorus.

Cortador (*koh-tah-doh*): 1. Cutter 2. The name for third *surdo* that is free to improvise.

Corte (*koh-chee*) or **Cortando** (*koh-tahn-doo*): The name for the third *surdo* improvisation that typically involves upbeats and variations alternating with the standard pattern.

Cuíca (*coo-ee-kah*): A friction drum played by squeezing and rubbing a wet cloth along a thin stick glued to the inside of the drum head, the *cuíca* is believed to have originated as a hunting tool in Africa.

D

Desenho (*gee-zehn-you*): A rhythmic design or theme, generally executed by the *ala dos tamborins*.

Desfile (*jish-feely*): Parade; *carnaval* parade.

E

Escola (*aysh-cola*): 1. School 2. The *carnaval* samba associations that are at the heart of the social life of the *favela*, consisting of a *bateria* of thousands, many ranks and types of dancers, choir, president, the front commission, the elders, Baianas (old ladies, or “aunts”), managerial hierarchy, costume designers, float designers, float pushers, and most of the rest of the neighborhood as well as paying outsiders. A top-rank *escola de samba* today numbers around 5000 just in performers.

F

Favela (*fah-vay-lah*): The slums of Brazilian cities, usually built precariously up hillsides.

Frigideira (*free-zhee-day-rah*): Kitchen implements used as samba instruments during *carnaval*.

Forró (*foh-hoh*): The family of styles and genres of northeast Brazilian cowboy music.

Frevo (*fray-voh*): A fast dance and music style from the northeast of Brazil, and one of few that are *not* a sub-genre of *forró*.

G

Gã (*gun* or *gao*): The single bell used in *Candomblé*.

Ganzá (*gahn-zah*): Tubular shaker with seeds inside ; quieter than a *chocalho*.

Ginga (*zheen-gah*): 1. A concept in *capoeira* that incorporates dynamism, awareness and balance. 2. The basic move in *capoeira*.

G.R.E.S. (Grêmio Recreativo Escola de Samba) (*greh-miu heck-rayah-chee-voo aysh-cola gee some-bah*): Recreational Samba School Association, the official title that precedes the name of each samba school that qualifies to take part in the *carnaval* parade and competition by having all the required elements.

L

Lê (*leh*): The smallest atabaque.

M

Maculelê (*mah-koo-leh-leh*): A type of martial art and dance performed with long bamboo sticks.

Mangue Beat (*mahn-gui beat*): A musical movement in the northeast of Brazil that combines local folkloric styles (such as *maracatú*, *forró*, and *côco*) with electronica, hip hop, and psychedelic and punk rock.

Maracatú (*mah-rah-kah-too*): Two styles of folkloric music from the northeastern Brazilian state of Pernambuco.

Mestre (*mesh-chee or mesh-tree*): Maestro, master; one who can play all or most of the instruments in the *bateria* at virtuoso level.

Mestre Sala (*mesh-tree sah-lah*): The male dancer who accompanies the *porta bandeira*.

Modinha (*moh-djeen-yah*): A style of Brazilian vocal music based on Portuguese *fado*, and primarily practiced in past centuries by the Portuguese.

Monobloco (*moh-noh-blow-coo*): A percussion/cavaco/vocal ensemble from Rio de Janeiro combining the samba-school sound with northeastern-Brazilian music, rock'n'roll and hip hop, Monobloco is among the first and rhythmically "sickest" of the move from carnival-only samba schools to touring, performing, recording bands in the samba world, perhaps influenced by the successful blocos of Bahia and the Mangue Beat movement of the northeast.

Morro (*mo-hoo*): 1. Hill 2. The *favela*.

O

Olodum (*Oh-low-doon*): World-famous samba-reggae band and bloco from Bahia. Olodum was founded by the developer of samba-reggae (from samba-afro and reggae), Neguinho do Samba, in 1979. Since then, they have recorded and performed in Central Park with Paul Simon, took part in a Michael Jackson video, and shared the stage with dozens of legendary artists ranging from Trilok Gurtu to Peter Tosh. Since 1984, Olodum is also a social institution working to improve life in the *favela* of Pelourinho. In this capacity, Olodum provides education in music, handicrafts, computer skills, and foreign language to the many abandoned or otherwise disadvantaged children of the city of Salvador da Bahia. Olodum finances these activities primarily with their own music and art work. The group is also a voice against racism in Brazil and around the world, and works to raise black awareness through their lyrics. Led by 26 directors, the group consists of hundreds of musicians and thousands of members, and thus can perform simultaneously in different parts of the world. They have more than a dozen full-length albums available internationally.

Olódùmarè (*Oh-low-doom-ah-ray*): The God of all gods: the creator in both *Candomblé* and the Yoruban (African) *Ifa* religion in south-eastern Nigeria and south-western Benin.

Orixá (*oh-ree-shah*): Deities representing various forces of nature, angels and ancestors in the Yoruban *Ifa* and Brazilian *Candomblé* religions, the *orixas* are associated with Catholic saints as well as colors, days of the week and types of food.

P

Pagode (*pah-go-gee*): A style of samba that initially emerged as low-key family-style backyard or restaurant-gathering entertainment centered around songs and accompanied by a cavaquinho, silverware, dishes and some quiet percussion, *pagode* has grown into a sophisticated form of studio and stage music with superstars (Zeca Pagodinho) and boy-bands (Revelação). The formalization of *pagode* is likely to have started with its primary songwriter, Jorge Aragão, and the super-group that invented some of the characteristic instruments of *pagode*, Fundo de Quintal (literally, the deep end of the backyard).

Palinha (*pah-leen-yah*): Call-and-response introduction to *batucada*, these days more often called a *paradinha* (even though it is a start, not a stop).

Pandeiro (*pun-day-roo*): A frame drum that looks like a tambourine, but isn't, the *pandeiro* is constructed so as to have minimal jingle volume. It is a hand drum played using a family of complex techniques which are at their most sophisticated in *chorinho* in order to allow the production of many tones from one drum. Using the combined manipulation of both hands, the *pandeiro* can reproduce the roles of all the drums in samba. The *pandeiro* is used in *pagode*, *axé* and many other styles of samba, except in the *batucada* and *samba de enredo*, where it cannot be heard.

The *pandeiro* is also used in *capoeira*, though the technique is significantly different than in samba. Low-tuned wood-framed drums of natural skin are preferred for *chorinho* and *capoeira*. Plastic-framed, tightly tuned drums with synthetic heads are used for *pagode* and *axé* (especially in a type of *axé* called *pagode baiana*), and large (12") synthetic pandeiros are used for acrobatics in the *carnaval*.

Paradinha (*pah-rah-djeen-yah*): Literally meaning "a small stop," this term has come to take the place in *palinha* in recent years, perhaps as *palinhas* have come to be performed more frequently and not just at the very beginning of a performance.

Partido Alto (*pah-chee-doo ow-too*): 1. One of the historic styles of music that contributed to the synthesis of samba from European and African elements. 2. One of the early names for samba. 3. One of the primary, fundamental rhythms of modern samba.

Passista (*puss-ish-tah*): A top-rank samba dancer; one of the most coveted positions in a *carnaval* parade.

Paulista (*pow-lees-tah*): A person from São Paulo.

Plata (*plah-tah* or *prah-tah*): 1. Jingle. 2. Cymbal.

Platinela (*plah-chee-nay-lah*): *Pandeiro* jingles.

Pratos (*prah-toos* or *prah-toosh*): Hand-held orchestral crash cymbals.

Primeira (*pree-may-rah*): First (song section or instrument type).

Primeiro (*pree-may-roo*): First (instrument type or role).

Puxador (*push-a-door*): Singer of the *samba de enredo*.

Q

Quadra (*quad-rah*): A large warehouse or similar space where the samba school of each *favela* practices and parties late into the night several times a week for nine or ten months before each *carnaval*. These practice/parties are where composers compete for the honor of having their song be the year's *carnaval* theme, dancers compete for the prestigious and few *passista* spots, percussionists simultaneously audition for and learn the *carnaval* material, and lead percussionists write the new designs. Thanks to visiting tourists, Brazilian or foreign, these practices are also a source of income for the community.

Quilombo (*key-lome-boo*) : City-states founded by escaped slaves (and Natives) during the centuries of slavery. It is believed by some that some *quilombos* still exist in the Amazon that have not had contact with the outside world and where the residents do not know that slavery is over. In some cases, such residents of recently discovered *quilombos* were taken advantage of by land-developers. The most famous *quilombo* was the Afro-democratic and long-lived state of *Palmares*, last led by the elected king Zumbi, who is said to have held off Portuguese attacks for many years. Zumbi is now celebrated as a national hero in Brazil.

R

Rebolo (*hay-bough-loo*): The middle of the three *Pagode* drums invented by Sereno of *Fundo de Quintal*, the *rebolo* has a metal or plastic, perfectly cylindrical body and one synthetic head covered with fake leather to give an unexpectedly deep, resonant tone. Played with one or both hands (one on the head and one on the body, which hand can be replaced by a brush), the *rebolo* plays the role of the third *surdo*.

Reco-Reco (*hay-coo-hay-coo*): A metal scraper, similar to the Cuban güiro, used in *pagode* and some *batucada*.

Refrão (*heff-roun*): Refrain; one of the three sections of the *samba de enredo*.

Regional (*hay-jyu-now*): The fast and show-oriented form of *capoeira*.

Repinicado (*hay-pee-knee-kah-doo*): The basic rhythm of samba; rolling samba 16ths.

Repinique (Repique) (*happy-neek or hepp-pee-key*): Known as the smallest of the *surdo* family, the *repique* is confusing to outsiders because it looks like a

drum-set tom-tom, sounds like Cuban *timbales*, and is played like a Senegalese *sabar* (with one hand and one stick in Rio-style samba). In *samba-reggae*, the *repique* is played with two long, flexible wooden sticks and is plays the combined role of the reggae guitar and the rumba *claves*.

Historically, a member of the *surdo* family, this is the high-pitched lead drum that makes the calls (signals all introductions, and resets the tempo at transitions). *Repiques* are also found among the ranks in the *bateria*, serving to keep the heavy *surdos* pushing forward (in time) during the parade. Although the repique sounds a lot like Cuban-style *timbales*, it lacks the low tones of *timbales*. However, the one-hand/one-stick method of playing (in Rio-style) allows for a lot of expressive variety.

Repique de Mão (*hepp-peek-ee gee moun*): Literally “repique of the hand,” this is a single-headed version of the repique, played with the hands. It is the leader and the smallest of the three *pagode* drums introduced by Sereno of *Fundo de Quintal*.

Roda (*hoe-dah*): Circle.

Rum (*hoon*): The largest (“mother”) of the three drums of the *Candomblé*, played only by ordained priests, the *rum* improvises within a wide but specific repertoire to accompany a dancer/worshipper who is “mounted” by an *orixa* (deity) during a ceremony.

Rumpri (*hoom-pee or hoom-pree*): The middle drum in *Candomblé*.

S

Samba (*some-bah*): The national music of Brazil, samba means several things at once: Rio-style samba, any music with a samba feel and samba beat, and any traditional music from Brazil. Samba is generally believed to have evolved through the synthesis of traditional music and dance forms from all over Africa (especially Nigeria, Benin, Angola, Mozambique, Sierra Leone, Burkina Faso, Guinea and Senegal) within the framework of European folk music, with possible early contributions from Brazilian Natives in the *quilombos*.

Samba-Reggae (*some-bah haggy*): A recent, popular synthesis of samba, *samba afro*, reggae and Dominican *merengue*, the term was first coined by Neguinho do Samba, who left the legendary *samba afro* group *Ilê Aiyê* to start *Olodum* and the *samba-reggae* movement. Today, due to the international success of *Olodum*, other *samba-reggae* acts like *Timbalada*, and the derivative pop music

called *axé*, the original *samba afro* form is now also referred to as *samba-reggae*.

Samba de Caboclo (*some-bah gee kah-boke-loo*): A type of folkloric samba developed in the *quilombos*, and played on *atabaques*.

Samba de Roda (*some-bah gee hoe-dah*): 1. A type of folkloric samba developed in the *quilombos*, and played on *atabaques*. 2. *Batucada* played in a circle as an improvisatory game.

Samba no Pê (*some-bah noo peh*): Literally “samba in the feet,” meaning the samba dance, or dancing the samba.

Sambas de Enredo (*some-bahs jean-hay-doo*): Samba songs from each samba school featuring the *carnaval* theme (*enredo*) of the year. Also, the style of *batucada*-with-song associated with *carnaval*.

Sambista (*some-beess-tah* or *some-beesh-tah*): 1. Samba dancer. 2. One who does samba.

Saudade (*sau-dah-gee*): A melancholy emotional element to the vocal and lyrical content of many samba songs, and a form of longing-based sadness, Portuguese-originated *saudade* is the opposite of Yoruba-originated *axé*.

Segunda (*say-goon-dah*): Second (song section or instrument type).

Segundo (*say-goon-doo*): Second (instrument type or role).

Subida (*soo-bee-dah*): “Getting on the bus.” *Subida* is a type of *crescendo* of swung triplets (or a more elaborate form of entrance design) performed by the *ala dos tamborins* at the beginning of each *samba primeira* or other relevant musical section.

Suinge (*swingy*): Swing, as in samba feel.

Surdo (*sooh-doo*): Literally the male form of the adjective ‘deaf’, a *surdo* is a bass drum, played with mallets, ranging from 12 to over 30 inches in diameter. The body is metal; the heads can be natural, synthetic (regular drum-set heads) or synthetic *napa* (fake leather) heads. In Rio-style samba, the *surdos* are tuned to two or three distinct tones and played with a single mallet, with the free hand used for muting. In samba-reggae, *surdos* are tuned to three, four, or five distinct tones, are built shallower and wider, and played with two mallets without muting.

Surdo Marcação (*sooh-doo mah-kah-soun*): The “marking” *surdo*, the heartbeat of samba, “*surdo de primeira*.” Olodum, however, calls all their *surdos* *marcação*: *marcação fundo*, *marcação 1*, *marcação 2*.

Surdo Mor (*sooh-doo moh*): The style and type of *surdo* that is unique to the Rio samba school *G.R.E.S. Estação Primeira de Mangueira*.

Surdo (de) Primeira/o (*sooh-doo pree-may-rah/roo*): First *surdo*; the lowest-tuned of all Rio-style drums.

Surdo de Resposta (*sooh-doo gee heh-poh-shtah* or *sooh-doo gee haysh-poh-shtah*): The *surdo* that responds, or “surdo segunda”.

Surdo (de) Segunda/o (*sooh-doo say-goon-dah/doo*): Second *surdo*, tuned to a pitch between the first and third *surdos*; plays the downbeat. The first official samba school of Rio De Janeiro, *G.R.E.S. Estação Primeira De Mangueira*, does not have the second *surdo*, and is immediately recognizable for it.

T

Talabarte (*tah-lah-bah-chee*): Strap.

Tamborim (*tumm-bo-rii*): 6” frame drum played in a variety of ways depending on the style, the tamborim was traditionally equipped with cat hide, but mainly uses synthetic heads today. In *batucada*, the tamborim is the loudest instrument in the *bateria*, tuned ear-shatteringly high and played with a type of whip. The *batucada*-style tamborim is possibly the most difficult samba percussion to play.

In samba jazz and *pagode*, the same drum is tuned down and played softly with a finger or wooden drum stick, achieving a completely different timbre. It is an extremely versatile instrument for being no more than a piece of plastic stretched across a round frame.

Tamborica (*tuhm-boo-hee-kah*): A tree-like mechanical device for playing ten tamborins at once, also known as a tamborim tree.

Tantã (Tan-Tan or Tan-Tão) (*tunn-tau or tunn-tunn*): The largest of the three *pagode* drums, the *tantã* is perfectly cylindrical with a metal body and *napa* (fake leather) head, and has an even deeper tone than the *rebolo*. Its role is to play the *batucada* part of the *surdo marcação* in *pagode*.

Tarol (*tah-roll*) ; plural: **taróis** (*tah-roysh*): See *caixa tarol*.

Teleco-Teco (*tell-eck-oo tech-oo*): A cental rhythm pattern (and its variations) in samba that indicates *partido-alto* direction without being the *partido-alto* rhythm.

Tercera (*teh-sarah*): Third (song section or instrument type)

Tercero (*teh-cerr-oo*): Third (instrument type or role).

Timbal (*chin-bau*): The popularized, mass-production form of atabaque, similar in shape and size to the African ashiko, but light enough (plastic) to be carried around during the Salvador *Carnaval*; used in *samba-reggae*, *samba afro*, and *axé*. Plural: **timbais** (*chin-baish*).

The naming of this drum is not consistent among Brazilians. Many call it *atabaque* and other names. This is not the Cuban *timbal* (pronounced *tim-bahl* in Spanish). It's also different from the small Brazilian *timba* (*chin-bah*) drum sometimes used in *pagode*, and from the Cuban popular music *Timba* (*tim-bah*). To make things worse, some companies will call the *timbal* “timba” instead, and may or may not include the smaller *pagode* drum in the same category.

Tremiterra (*tray-me-teh-huh*): Literally “earthquake,” the largest *surdo* with a diameter of over 30 inches.

Triângulo (*tree-ahn-glue*): Triangle (not isosceles like the orchestral triangle, but with three unequal sides).

Trio Elétrico (*tree-oh eh-let-rick-oo*): 1. A trio of *puxador* (singer), *surdo* player and *cavaco* player and PA system on the back of a flat-bed truck at *carnaval*, leading the drummers and revellers. 2. The truck carrying the trio.

V

Velha Guarda (*vell-yah goo-ahr-dah*): The elders (“Old Guard”) of the samba school.

Violão (*vee-oh-loun*): Nylon-string guitar (*not* a violin).

Virado (*vee-rah-doo*) or **Virando** (*vee-run-doo*): 1. Literally “turn,” a punctuating rhythmic figure performed by the tamborins and caixas at phrase ends. 2. *Third-surdo* improvisational lick featuring a long series of offbeats (and only offbeats), also known as *prafrente* [2].

X

Xequerê (*sheh-keh-reh*): Also known as cabasa, but not to be confused with the LP afuche/cabasa, a hollowed-out, decorated gourd, played by shaking; also known as *xequerê*, which is pronounced the same as the more familiar Cuban *shekere*.

Z

Zabumba (*zah-boom-bah*): Shallow (thin), metal-shelled bass drum used in *forró*, played with one stick and one mallet.

F.2 Clave Terminology throughout the Diaspora

What follows is an inventory of terms given to the clave concept (clave direction and related notions) by scholars, teachers, and musicians around the world:

- “The principle of mobility and finality,” W. Anku, African musicologist, [3]
- “Regulative rhythm pattern,” L. Ekwueme, Nigerian professor of music, [4],
- “*Topos*,” meaning “repeated theme,” K. Agawu, Ghanaian musicologist, [5],
- “African hemiola,” R. Brandel, [6],
- “Time line,” Professor J.H.Kwabena Nketia [7]
- “Standard pattern” and “time line” Agawu, K., [8]
- “Clips,” a type of hand-clapped clave-like rhythm in Ghana
- “Asymmetrically arranged syntactic division,”
- “Disjunct cyclic organization,” List [9],
- “Pan-African standard pattern,” A.M. Jones, as reported in [10].

F. 3 Musicological Definitions, Brazilian, European and Global

The arguments in this section are informed by the author’s observations and supported by numerous written sources, as well as interviews¹⁷² with a small number of

¹⁷² The researcher has maintained human-subjects certification during 2003, 2004, 2005, and 2006, as part of his work for the university’s Assessment Initiative. The time period for the interviews falls within this interval.

professional performers and scholars of music, bolstered by testimonials from a cross-section of the Portland, Oregon Brazilian music scene.

One of the informants is master *Candomblé* musician Jorge Alabê¹⁷³, possibly the highest ranking (non-Jazz) samba drummer in the United States. He is the author's teacher, the teacher of his teachers, and the teacher of his teachers' teachers. In turn, the author's humble qualification as a performer and teacher of Brazilian music is attested to by the times he has accompanied the Alabê either as an assistant or as a drummer, in workshops, classes and performances.

F.3.1 Definitions of Music

A definition of music that is both precise and broad, and that everyone can agree on has not been found, but several attempts come close to capturing the essence of what music might mean to a majority of those involved with it.

There seems to be general agreement that music is sound intentionally organized by human beings. As early as 1880, Elson quoted another author (not properly cited) as having defined music as “the art of moving the feelings by combinations of sounds” [11], a definition that could be quite agreeable to today's avant-garde, gothic rock, industrial rock, or metal listener, unlike much of the rest of Elson's book, in which India, China, Japan and “Africa” are presented as “savage” nations.

The New American Dictionary of Music defines music as “sound organized in space and time” [12]. According to Blacking, “Music is sound that is organized into socially-accepted patterns...” [13]. Differentiating between *music*, *Music*, and *MUSIC*, Elliott states: “MUSIC is a diverse human practice consisting in many different musical practices or Musics. The word music (lowercase) refers to the audible sound events, works, or listenables that eventuate from the efforts of musical practitioners in the contexts of particular practices.” [14] In criticizing and expanding on Elliott's writings, Panaiotidi says “MUSIC is the diverse human practice of overtly and covertly constructing aural-temporal patterns for the primary (but not necessarily the exclusive) values of enjoyment, self-growth, and self-knowledge.” [15]

From these progressively broadening definitions, one can take away some key factors common to all definitions of music: sound, organization, time, context, and

¹⁷³ An Alabê is the highest-level spiritual leader in the Afro-Brazilian religion of the Ketu nation of *Candomblé*, a master drummer who ranks higher than the highest priest. One only needs to hear Jorge play any samba instrument to appreciate the depth of his expertise. Like a 3-D drawing of a cube on a page, one can hear several patterns when Jorge plays any one pattern. He does so with incessant drive and energy, and perfect swing. Even when he is playing just the *repinicado*, one can hear the lilt of *partido-alto* and the accents of *teleco-teco* in the subtle dynamics of his phrasing. That's a *bamba*.

practitioner. In a much more succinct and much less precise attempt, Kerman defines music simply as “the art of sound in time.” While this is a sensible and practically useful definition, it is heavily dependent on a definition of art, which is even harder to come by. The author would like to propose here yet another definition of music that he believes to be sufficiently general and precise:

Music is sound, intentionally arranged in time through some human involvement to constitute a formal or informal entity.

The provision for formal and informal arrangements of sound could allow for such acts as whistling in the shower to be considered music. Similarly, the use of the term ‘arranged’ rather than ‘organized’ makes the definition inclusive of Musique Concrete, Found Sound, the more eccentric examples of No Wave and so-called Noise Music, and even nature recordings.

F.3.2 The Elements of Music

As stated previously, the small number of musicians and music scholars surveyed about what music is made up of was a convenience sample. Nonetheless, a degree of diversity was achieved, as the interviewees included formally trained scholars, informally and formally trained professional performers, and informally trained educators, from both North and South America. In particular, they were a classical pianist and music professor from the Music Department at Portland State University, a Brazilian master drummer, two professional musicians from the United States (one a member of an internationally successful pop group, the other, an accomplished practitioner of authentic Cuban, Brazilian and Nigerian music), and a group of four professional music educator/performers, who provided informal but valuable insights. Any subsequent references to the interviewees will be as informants 1, 2, 3, 4, and the informant group, respectively. The ratio of informants with significant northern music-theoretical training to the total is 50%.

The Elements of Music, according to the Selected Informants

According to Informant 4, the elements of music are “tension and release.” According to Informant 1, they are “rhythm, melody, and harmony” with rhythm as the most important aspect connecting melody and harmony. According to Informant 2, the four elements that make a real musician are “*suinge* (swing), *jinga*, *cadência* and *malemolência*.” According to some members of the informal-informant group, music is made of “sound and organization,” or again, “rhythm, melody, and harmony.”

In conceiving this article and setting out to conduct interviews, the author also started out thinking that the elements of music were rhythm, melody, and harmony. One immediate conclusion from this is that the author's view of music is very much in agreement with the tradition of northern-based Art Music. If this can be taken as an indication of some small degree of northern bias in the author's approach to music, let it be noted as such.

The written sources consulted also have a similar take on what elements constitute music. Even though the terminology used for such elements also involves concepts like 'parameters', 'properties', and 'canons', with one significant exception there is a consensus that rhythm, melody, and harmony are the primary building blocks of music.

Likewise, in attempting to help young people cultivate a sense of good music, Henderson declared in 1925 that "the material of form in music consists of rhythm, melody, and harmony." [16]

Though they disagree with much that Henderson says in his book on cultivating taste, Nettl [17] and Hood [18] also seem to think that the elements of music are rhythm, melody, and harmony, along with form and texture, as indicated by the following: "... each example could also be examined by a set of categorical parameters, such as rhythm, melody, harmony, form, texture, and all the other canons in which scholars of Western music place their analytical faith." [18]

Similarly, Riddley repeatedly refers to music as having the primary properties of "melody, harmony, rhythm" which he calls basic perceptual properties [19].

Even though many other properties of music can be listed, such as texture, counterpoint, production, tonality, modality, form, modulation, clave, polyrhythm, lyrics, contour, and timbre [20], all of these are manifestations of rhythm, melody and harmony.

While rhythm, melody, and harmony appear to be sufficient and reasonable candidates for the basic elements of music, a brief look at their definitions would suggest that though rhythm has to do with time, and is therefore a *horizontal* property, melody and harmony involve both *vertical* (simultaneous) and *horizontal* (temporal) aspects.

Melody is defined in The New Everyman Dictionary of Music as "an intelligible succession of notes defined by pitch and rhythm. [Furthermore, in] western music it's unusual to find melodies which do not at least imply harmony." [21] Similarly, according to Kerman, melody is a "coherent succession of pitches played or sung in a certain rhythm" [22, p. 20]. In other words, this sequential pitch relationship of musical notes (sound events) cannot exist without rhythm.

Similarly, harmony is defined as "the musical effect derived from combining different pitches simultaneously" [21]. Again, according to Kerman, harmony is "the

relationship of tones considered as they sound simultaneously and the way such relationships are organized in time.” [22, p. 20] Thus, we see that harmony is also a function of time, and it is both a horizontal and a vertical property.

In the mathematical sciences, orthogonal bases are used to represent the fundamental elements of any *vector space*, which can be thought of as the domain of a set of relationships between measurable quantities. The interaction of these quantities generates combinations of values with distinct characteristics, each represented by a different point in space, in other words, a vector. A set of orthogonal bases are the defining components of a space such that any object in the space can be expressed as a combination of those bases, and that no single basis can be expressed in terms of any of the others.

Considering music as a mathematical *space* defined by its elements (where the two dimensions are frequency and time), the musical equivalent of orthogonal bases would be an exhaustive set of elements that can be defined without reference to one another. Since the definition of melody requires reference to rhythm, and the definition of harmony requires reference to melody, only rhythm is a candidate for an orthogonal basis. This is in agreement with the view, expressed by Informant 2, that rhythm is the single most fundamental element in music. *Pitch* is what we add to rhythm to create ‘melody’, and to make ‘harmony’ possible. Thus, it may be sensible to argue that a sufficient basis set for the space of music (e.g., a vector space used to represent music) may consist of only two elements: pitch and rhythm. Indeed, Kerman argues a similar set of two elements:

[I]t may be helpful to think of pitch and time as the two main dimensions or “coordinates” of music. A graph with pitch reading up and down on the vertical axis, and time running from left to right on the horizontal axis, can help in the conceptualization of music In fact, such a pitch/time graph comes quite close to musical notation. [22, p. 23]

‘Frequency’ may be an even better term, since ‘pitch’ is dependent on ‘loudness’ (related to amplitude) and ‘timbre’ (the mix of partials, or harmonics). Thus we have the orthogonal bases of music as ‘frequency’ and ‘time’. (We may perhaps add ‘amplitude’ to that set for music as it is performed, resulting in a three orthogonal bases, but frequency and time are sufficient for conceptual analysis.) Not so coincidentally, this is the view of music from the point of engineering; from Fourier analysis to Wavelets, musical signals are analyzed as frequency components and their magnitudes evolving over time.

In order to gain further musical understanding of these concepts—especially the relationship between ‘rhythm’ and ‘time’—let us take a closer look at the definition of rhythm and related concepts.

F.3.3 Time, Rhythm and Meter

Although simplistic, Henderson’s definition that “rhythm is the grouping of sounds with reference to their duration and accent” [16] points out two important parameters of rhythm: *time* (duration) and *dynamics* (accent). Since dynamics can be considered a vertical aspect, *time* emerges as a better candidate than *rhythm* for an orthogonal basis.

According to The New Everyman Dictionary of Music, rhythm has to do with all aspects of music involving time: “In its largest sense the word means all that is concerned in music with matters dependent on time, such as the metre, [...] the distribution and balance of phrases, etc.” Kerman agrees: “Rhythm, in the most general sense, is the term referring to the whole time aspect of music.” [22, p. 18]

Aspects of rhythm include such elements as *tatum*, *beat*, *pulse*, *swing*, *rubato*, *meter*, *tempo* and *syncopation*. All of these definitions are interdependent when put into words, although most of them are natural to musicians. Simple definitions for some of these could be given as in the following: *Tempo* is the time rate of the occurrence of beats, where a *beat* is the fundamental timing unit in metrical music, in which the *meter* indicates the accent structure. Meter is the metronomic sequence of temporal subdivisions implied by a piece of music, as commonly felt by listeners. In Kerman’s words, meter is “any recurring pattern of strong and weak beats” [22, p. 19].

Pulse is sometimes thought of as the explicit statement of the beat or of some multiple of the beat, usually on a percussive instrument with little or no sustain. *Tatum* is the smallest interval of time implied or indicated by the totality of the note events in a piece of music with steady tempo, with reasonable allowance for musically-acceptable tempo and timing variations and rhythmic embellishments such as *flams*¹⁷⁴. This is closely related to, if not the same as, Kwabena Nketia’s term “density referent” [23, p. 396] and Waterman’s “metronome sense” as reported by Merriam [24, p. 14].

It is the passing of a uniform, abstract time unit that (northern-trained) drummers measure. Thus, engineers, physicists and mathematicians are the closest in their thinking to northern-trained drummers.

¹⁷⁴ To put it mathematically, it is the lowest common denominator of rhythmic timing, an imagined uniform beat series of frequency equal to (when indicated), or the smallest necessary integer multiple of (when implied), the highest-frequency component of an audio signal.

Many dancers (in general) and some Latin American drummers count *events* (as opposed to rests, non-events, or absolute intervals) when they numerically express rhythms. Some master Latin American percussionists are known to count the 3-2 *son* clave by saying: “1·2·3·4·5·” (where “·” represents the length of time corresponding to the passing of one silent tatum). Similarly, dancers, and occasionally even choreographers and dance teachers will refer to a rhythm as being “in three,” which they proceed to spell out as “one, two, three, (rest); one, two, three, (rest).” In contrast, a northern-trained drummer would refer to this as being “in four” because of the passing of uniform time units, regardless of whether an event (a dance step, a note onset, etc.) occurs or not.

The smallest of such *theoretical* time units is called the *tactus*. The smallest of such *practical* units, determined by the temporal interaction of all notes played, is called the *tatum*. The difference arises from the fact that humans do not execute musical timing in the same way that it is written or otherwise codified for non-aural communication. Writing music down, whether in European notation or not, is an abstraction. The degree of departure from actual execution varies based on the tradition of performance and on the method, tradition, and precision of transcription. While a small number of instances of Baroque music and certain pieces of electronic music may be meant to be performed strictly as written, most folkloric, popular, and even orchestral music is performed with a characteristic deviation from the grid-like mechanical timing implied by the necessary strictness of transcription. This is variedly known as *feel*, *expression*, *swing*, *soul*, and a variety of other names. (Contrary to popular opinion, swing is not limited to jazz and blues. Almost every type of music, from Scandinavian *Hardingfele* and Japanese Shinto ceremonial music, to Senegalese *Sabar Wolof* and Greek *Rebetika* has its own style of expressive timing.) The accurate representation of such characteristics in transcription is not only not useful, but possibly detrimental or even impossible—the former because such scores would be impossible to sight-read, and the latter because the extent of timing variation can differ from performer to performer even within a given tradition. The same information can instead be loosely stated via stylistic markings or comments, and would be readily understood by experienced musicians.

F.3.4 Brazilian Musical Definitions from Jorge Alabê

From a conversation with Jorge Alabê, master drummer/singer/dancer/high priest (Alabê) of the Ketu Nation of Afro-Brazilian *Candomblé*.

Jinga is how one weaves in and out of situations, whether in life, music, fighting, dancing, or anything else. Some folkloric samba builds and maintains interest via entrances and exits. A proficient or expert sambista would know how to interpret

the playing of other musicians to take part in this collective sculpting, much as is done in jazz.

Cadença is what American jazz, pop and samba musicians call ‘time’. To have internal time, and to follow the time of others impeccably come foremost among the elements that make up an expert samba musician.

Malemolência is the way a carioca (a native of Rio) approaches samba in particular, and life in general. It is an attitude of laid-back confidence¹⁷⁵. One way to cultivate this attitude is described in the book *Samba* by A. Guillermprieto, in the section on the men’s samba dance [24].

Suinge (swing) is the soul of the rhythm. In technical terms, it is the very human and very elusive variations in micro-timing and micro-dynamics that are associated with a given tradition. It is also called expressive timing. While there are those who claim all swing outside of Jazz is simply dynamics (accenting differences), there is evidence that swing exists in all music in different ways and to varying extents. Not only is swing not limited to Jazz, it is also not limited to musics of the African Diaspora. The author has heard from a Finnish-Swedish violinist that Scandinavian folk music has an entirely different feel when played by authentic folkloric musicians and when played by others sight-reading a score of the same piece. Whether that is a result purely of accenting or whether timing variations play a role is beyond the realm of this article. However, it is well within the concerns addressed here that *suinge* in samba is a result of both differential accenting and expressive timing.

Expressive microtiming and *suinge* in Brazilian music is an emerging area of research in Computational Ethnomusicology ([25, 26, and 27]).

¹⁷⁵ Jorge Alabê is a *carioca*, and this is his definition. The term can also be found in the song *Menino do Pelô*, composed by Saul Barbosa & Gerônimo, about the Pelourinho neighborhood of Salvador da Bahia.

References for Appendix F

- [1] Castro, R., *Rio de Janeiro: Carnival under Fire*, [translated by John Gledson from the original *Carnaval no fogo*, Companhia das Letras, Brazil], London, UK: Bloomsbury Publishing Plc., 2004.
- [2] Mestre Jonas via Dr. Phil Galinsky, workshop, California Brazil Camp, Sep. 3, 2011.
- [3] Anku, W., "Towards a Cross-Cultural Theory of Rhythm in African Drumming," *Intercultural Music*, ed. Kimberlin, C. T., and Euba, A, Bayreuth, Germany: Breitinger, Vol. 1, 1995, pp. 167–202.
- [4] Ekwueme, L. E. N., "Concepts of African Musical Theory," *Journal of Black Studies*, Vol. 5, No. 1, 1974, p. 45.
- [5] Agawu, K., "Structural Analysis or Cultural Analysis? Competing Perspectives on the "Standard Pattern" of West African Rhythm," *Journal of the American Musicological Society*, Vol. 59, No. 1, 2006, pp. 1–46.
- [6] Brandel, R., "The African Hemiola Style," *Ethnomusicology*, Vol. 3, No. 3, 1959, p. 141.
- [7] Kwabena Nketia, J. H., "African Music and Western Praxis: A Review of Western Perspectives on African Musicology," *Canadian Journal of African Studies*, Vol. 20, No. 1, 1986, pp. 36–56.
- [8] Agawu, K., "Structural Analysis or Cultural Analysis? Competing Perspectives on the "Standard Pattern" of West African Rhythm," *Journal of the American Musicological Society*, Vol. 59, No. 1, 2006, pp. 1–46.
- [9] List, G., "African Influences in the Rhythmic and Metric Organization of Colombian Costeño Folksong and Folk Music," *Latin American Music Review / Revista de Música Latinoamericana*, Vol. 1, No. 1, 1980, p. 6–17.
- [10] Kauffman, R., "African Rhythm: A Reassessment," *Ethnomusicology*, Vol. 24, No. 3, 1980, pp. 393–415.
- [11] Elson, L. C., *Curiosities of Music; a collection of facts not generally known, regarding the music of Ancient and Savage Nations*, Boston: Oliver Ditson & Co., 1880, p. 7.
- [12] Morehead, P. D., ed., *The New American Dictionary of Music*, New York, NY: Dutton, 1991, p. 359.
- [13] Blacking, J., "The Value of Music in Human Experience," *Yearbook of the International Folk Music Council*, Vol. 1, 1969, p. 36–37.
- [14] Elliott, D. J., *Music Matters*, New York, NY: Oxford University Press, 1995, pp. 23–125.
- [15] Panaiotidi, E., "What Is Music-Aesthetic Experience versus Musical Practice," *Philosophy of Music Education Review*, Vol. 11, No. 1, 2003, p. 71–89.

- [16] Henderson, W. J., *What Is Good Music? Suggestions to Persons Desiring to Cultivate a Taste in Musical Art*, New York, NY: Charles Scribner's Sons, 1925, p. 18.
- [17] Nettl, B., "Types of Tradition and Transmission," *Cross-Cultural Perspectives on Music*, Falck, R., Rice, T., and Kolinski, M, eds., Toronto: University of Toronto Press, 1982, pp. 5–6.
- [18] Hood, M., "All Musical Cultures Are About Equally Complex," in *More Than Drumming - Essays on African and Afro-Latin American Music and Musicians*, Jackson, I. V., ed., Westport, CT: Greenwood Press, 1985.
- [19] Riddley, A., *Music, Value, and the Passions*, Ithaca, NY: Cornell University Press, 1995, p. 56–64.
- [20] Green, L., "Evaluation and assessment in the classroom," in *Evaluating Creativity - Making and learning by young people*, Sefton-Green, J., and Sinker, R., eds., London: Routledge, 2000, p. 98.
- [21] Bloom, E., ed. *The New Everyman Dictionary of Music*, New York, NY: Weidenfeld and Nicholson, 1988.
- [22] Kerman, J., *LISTEN: Brief Edition*, New York, NY: Worth Publishers, Inc., 1987.
- [23] Nketia, J. H. Kwabena, *African Music in Ghana*, Boston, MA: Northwestern University Press, 1970.
- [23] Merriam, A. P., "Characteristics of African Music," *Journal of the International Folk Music Council*, Vol. 11, 1959.
- [24] Guillermprieto, A., *Samba*, New York, NY: Vintage, 1991, pp. 97–99.
- [25] Wright, M., and Berdahl, E., "Towards Machine Learning of Expressive Microtiming in Brazilian Drumming." *Proceedings of the 2006 International Computer Music Conference*, New Orleans, LA, 2006, pp. 572-575.
- [26] Lindsay, K., and Nordquist, P., "Pulse and Swing: Quantitative Analysis of Hierarchical Structure in Swing Rhythm," *Journal of the Acoustical Society of America*, Vol. 122, No. 5, 2007, p. 2945.
- [27] Naveda, L., Gouyon, F., Guedes, C., and Leman, M., "Multidimensional microtiming in Samba music," In Pimenta, M., Keller, D., Faria, R., Queiroz, M., Ramalho, G., and Cabral, G., editors, *12th Brazilian Symposium on Computer Music* (Best Paper Award), 2009, pp. 1–12.

APPENDIX G: Northern/Non-Northern vs. Western/Non-Western

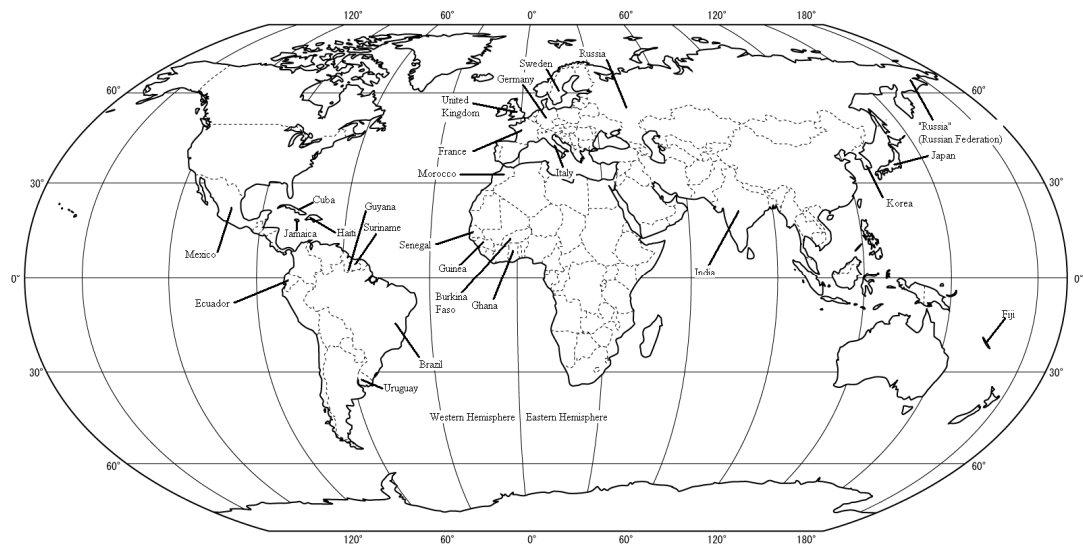


Figure 73: World map.

A quick look at a world map would confirm that countries such as Cuba, Brazil, Morocco and Senegal are located to the *west* of European countries like Greece, Russia, and Denmark. However, members of the latter group are counted among the *Western* nations, while the former are strangely dubbed *non-Western*. Upon closer scrutiny, a more accurate pattern emerges that has to do with north and south, rather than east and west.

Eminent musicologist and ethnomusicologist Nettl has interestingly titled a co-authored book containing chapters on Latin American music *Folk and traditional music of the Western continents* (Nettl, B., with chapters on Latin America by Behague, G.).

The classic examples of Art Music predominantly come from Germany, Austria, France, Italy, and Russia. These are called “Western” nations, perhaps because they are in the western portion of the world that Europe knew about (the “Old World”¹⁷⁶), and found relevant prior to the Iberian discovery of the Americas.

Some may argue that Russia is not considered “Western.” This, however, is not true in the arts or the sciences: Both in music (consider Mussorgsky, Rimsky-Korsakov, Glinka, Scriabin, Tchaikovsky, Prokofiev, Rachmaninoff, Shostakovich, and Stravinsky.), and in mathematics (Chebyshev, Markov, Anosov, Lyapunov, Kolmogorov, etc.), Russia has been a central player in “Western Civilization.”

¹⁷⁶ “Our conventional view of the ‘Old World’ as having three continents—Europe, Asia, and Africa—breaks down once we think about their physical and cultural geography.” (pp. 6–7, *A Concise History of the Middle East*, Arthur Goldschmidt, Jr., 2002, Cambridge, MA: Westview Press)

Goldschmidt argues that “the lands *south* [emphasis added] and east of the Mediterranean were *the East* [emphasis in the original] to our cultural forebears, until they started going to India and China” (p. 7).

As a result of this historical point of view, it is quite common to find Cuba, Brazil, and Morocco, for example, lumped together with Iran, Japan, and Fiji as “non-Western,” while Greece, Russia, and the Netherlands, different as they may be, are all referred to as “Western.”

We can see that the so-called Western nations are not necessarily located west of the non-Western nations. In fact, with a few exceptions, the so-called Western nations are all *north* of the non-Western nations. To be specific, the northernmost countries along the line delineating cultural “west” from “non-west,” which are Mexico, Cuba, the Bahamas, Morocco, Tunisia, Turkey, Armenia, Kazakhstan, China, Mongolia, North Korea, and a few others, are almost all *due south* of the southernmost “Western” nations: the US, Spain, Italy, Malta, Greece, and Russia. While a few from the non-northern list (Turkey and Korea) happen to be somewhat farther north than a few on the northern list (Greece and Spain), this lack of perfect delineation does not detract from the point that for the most part, what westerners and non-westerners alike have been thinking of as “Western” has not had as much to do with east/west as with north/south.

Furthermore, the official definition of Eastern Hemisphere and Western Hemisphere (separated by the Greenwich Meridian) are such that unmistakably “Western” nations like Germany, Switzerland, and Norway are located entirely in the Eastern Hemisphere, while typically the non-western Morocco, Uruguay, Guyana, Suriname, Haiti, and Senegal, to name a few, are all fully in the Western Hemisphere.

Hence, as a more geographically accurate expression, the use of “northern” in place of “Western,” and “non-northern” in place of “non-Western” is followed here.

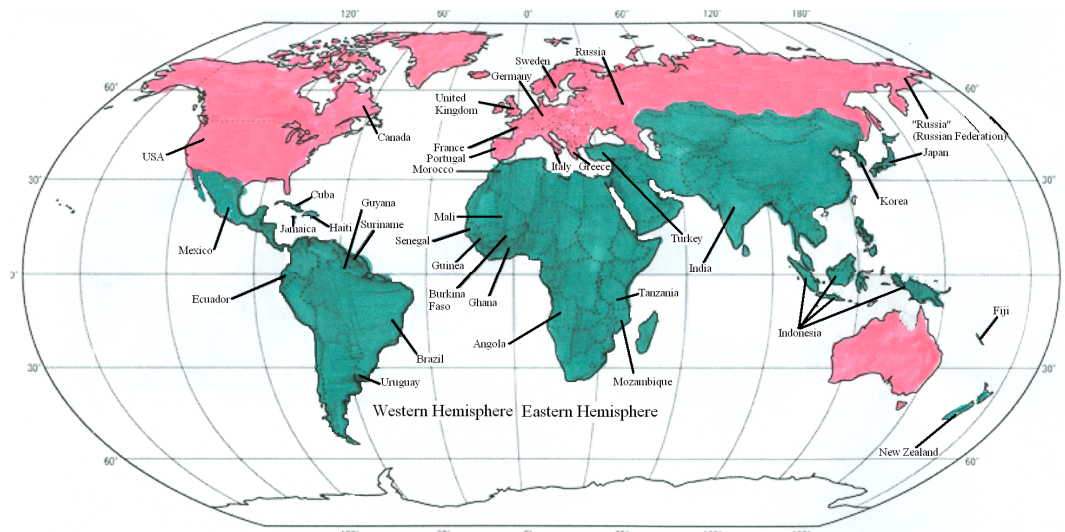


Figure 74: World map with the hemispheres and several countries relevant to the discussion.
The pink (lighter-colored) areas are (approximately) the extents of nations typically considered “Western,” and the green (darker-colored) areas are (approximately) the extents of nations typically considered “non-Western.”

APPENDIX H: Occurrences of Clave-Type Patterns in Music around the World, and Recommended Listening

This appendix lists examples of music that either displays the clave-direction concept or happens to have clave-like patterns (i.e., patterns that happen to have the clave concept expressed in their structure, intentionally or otherwise).

Section H. 3 speaks to the situation where one asks “How would this fit with the samba? If I wanted to steal this lick and throw it into a samba song, where would I place it so that I can repeat it a few times and not break the sambaness of the samba?” Such questions are valid because the Brazilians who are coming up with *tamborim* designs or *surdo-caixa bossas* for next year’s *carnaval* are also listening to 50 Cent, nine inch nails, and Queen.

H.1 Examples of Clave Direction in Brazilian Musics

“Capa de Revista” by Fundo de Quintal (pagode pioneers) from the album *Papo de Samba* is one of their many typical 3-2 sambas.

“Chapa Quente” by Fundo de Quintal from the album *Papo de Samba* is another example of samba in 3-2 *partido-alto* clave direction, but the *pandeiro* plays fewer of the clues in *partido-alto*, possibly because there seems to be one *pandeiro* filling the roles of two until just under a minute in when the second *pandeiro* and the *surdo* enter.

“Cinema Novo” by Caetano e Gil from the album *Tropicália 2* is an atypical Caetano Veloso song in that it is a straight-forward samba. (Caetano does not play much samba, though elements are found in some of his songs.) Though the percussion section builds slowly into a samba-school sound, the *partido-alto* clave direction is established immediately by the timing of the guitar chords (with respect to the vocals).

“Ela Não Gosta de Mim” by Agêpê from the compilation *O Samba: Brazil Classics 2* starts out with a layering of samba percussion (*agogô*, *pandeiro*, *tamborim*, *cuica*, and the clave-neutral *surdo*) hinting at bits and pieces of the *partido-alto*, along with the essential samba string instrument, the *cavaco*. But when the singer comes in the *agogô* and *pandeiro* do a little “hickup” and adjust to the new clave/*partido-alto* direction. Careful listening at high volume to the very beginning will also lead to the discovery of a low-tuned *pandeiro* far in the background, laying down the full *partido-alto*, presumably as a reference to the sparse *agogô* pattern that starts the song.

“Caxambu” by Almir Guineto is a fast *samba de roda* augmented by pagode and studio instruments (*cavaco* and drum set, respectively), also from the compilation *O Samba: Brazil Classics 2*. The clave-neutral hand claps can be heard during the line “é na palma da mão” but the rest of the song is driven by the 3-2 strumming on the *cavaco*. Furthermore, the *atabaque* (conga-like ceremonial hand drum) part shows one of the other functions of clave: phrasing without relative offbeatness.

“É Preciso Muito Amor” by Chico Da Silva, also from the compilation *O Samba: Brazil Classics 2* establishes its clave sense through the guitar and the *cavaco*.

“De Um Sinal” by Pimenta N’Ativa from the compilation *Axé Bahia 2000* is an *axé* (genre) song in the *pagode baiana* style. Throughout the song, the horns play exactly the same onset (rhythm) pattern as the *third surdo* plays in much of Rio-style samba *batucada*. (*Axé* music is not from Rio, but from the northeastern state of Bahia.)

“Posso até me apaixonar (acústico MTV)” by Zeca Pagodinho from the album *acústico MTV* starts out with a flute intro just barely in 2-3. After the turnaround and break about 10 seconds into the piece, the intro gives way to the body of the song where the *pandeiros* indicate *partido-alto*. In this and in the Fundo de Quintal examples (as is typical in *samba pagode*), there are two *pandeiros*, one playing *partido-alto* and one supplying the samba swing. The set of snare accents during the refrain is clave-neutral.

“Eu Sou Favela” by Bezerra Da Silva from the compilation *Favela chic: postonove* establishes clave direction on the first note, which, with the entrance of other instruments, is revealed to be a 16th after the downbeat of the 4. This clave direction is later confirmed by the light *tamborins* and the *cavaco* that enter around 28 seconds into the song.

Two songs from the album *Feijão Com Arroz* by Daniela Mercury are recommended for different reasons. The *samba-reggae* “Nobre Vagabundo” starts out with an *a capella* section where the vocals indicate 3-2 clave in extreme clarity.

Another song from *Feijão Com Arroz*, “Vide Gal” is interesting not only because it is (like “Cinema Novo”) a straight-ahead samba from the state of Bahia, but also because it adds horns (not allowed in Rio carnival) to the full samba-school sound. The horns and electric bass adhere to the clave direction of the samba school.

“O Mais Belo Dos Belos (O Charme Da Liberdade)” by Ilê Aiyê from the album *25 Anos* is a *samba-afro* in 2-3. This is most clearly heard in the sparse *repique* accents. At 37 seconds into the song, the heavy “kitchen” comes in. This can be considered flipping the clave direction to 3-2 by way of a half-phrase turnaround because after this point, the song feels like phrases start on the 3-side (although the percussion did not change its course).

“Por Amor Ao Ilê” by Daniela Mercury from the album *Musica de Rua* is a *samba-reggae* / *axé* where the main sense of clave direction is carried in the *repiques*. In fact, the pattern utilized here forms one of the main arguments of clave direction put forth in this dissertation (Figure 9, middle row, Appendix A).

“É Ruim de Quebrar” by Fundo de Quintal from the album *Papo de Samba* establishes *partido-alto* by 6 seconds into the song when the *cavaco* is done with its intro pattern and launches into the standard *cavaco* version of *partido-alto* in 3-2.

“O Dia Em Que Faremos Contato” by Lenine (*Brasil 2 mil*, 1999): The sparse guitar chords on the *e* after the downbeat of 1, and the *e*[~] after the downbeat of 3 indicate the clave to be 3-2.

“Choro de Lera” by Cascabulho from the album *É Caco De Vidro Puro* features intro percussion (as well as the rest of the song) in 2-3 clave direction. This song is an example of a northeastern-Brazilian style called *mangue beat* (pron. like monkey with a ‘g’) that combines *maracatú*, *frevo*, *coco*, *forró*, rock, samba, and electronica.

“Cavalo da Simpatia” by Carlinhos Brown from the album *Bahia Do Mundo* demonstrates onbeat/offbeat equivalence based on tactus reference. (See discussion on syncopation, offbeatness and the tactus, cf. Figure 14, bottom, of Appendix A).

“Tropicadelica (with Natalia Clavier)” by Ursula 1000 from the album *Mundo Beyondo* (2011) has vocals in 2-3 corresponding to the attack-point vector 1010100101010000, which is a truncated version of a very typical 2-3 pattern found in both Brazilian and Cuban music.

“Profissão MC” by Marcelo D2 (*Acústico MTV*, 2004): 2-3 *montuno* (Cuban-style).

“To Jackie With Love” by Marty Most, Jazz Poet from the album *Marty Most, Jazz Poet, presents DRUMSCUSSION (New Orleans–Brazil)* (2001) featuring Jorge Alabê and Roy Harper includes a prominent Brazilian bell pattern (in its 2-3 form here) that is identical to Cuban rumba *cáscara*. The same pattern is used in the same way in ...

H.2 Accessible Examples of Clave in other African-based Music, and Jazz

“1 nite stand” by John McLaughlin Trio (*qué alegría*, 1991): Main theme (guitar) in 3-2.

“Tracie” by The Headhunters [written by Bill Summers] (*Platinum*, 2011 [Owl Studios OWL00139]): congas, bass, trumpet, timbales, and the main theme on the horns (the “head”) are all in 2-3 clave direction.

“Tell Everybody” by Herbie Hancock (*Feets Don’t Fail Me Now*, 1979): prominent 2-3 *agogô* (using an old-fashioned samba pattern) performed by Brian Davis of Lions of Batucada and Pink Martini.

“Ugh!” by Art Blakey And The Jazz Messengers (*Midnight Session*, 1998): This album features numerous drum-set patterns that carry a clave sense, and the track “Ugh!” is a veritable catalog of 2-3 patterns, although Blakey’s presentation is more “exciting” and “artistic” than folklorically representative. Hence, multiple interpretations of the clave direction(s) of these patterns are possible (depending on the choice of reference).

“Casino” by Art Blakey And The Jazz Messengers (*Midnight Session*, 1998) is another piece where several patterns exhibit a sense of clave direction, though not necessarily by referencing commonly recognized patterns.

“Potpourri” by Art Blakey And The Jazz Messengers (*Midnight Session*, 1998): 2-3 clave direction manifested in the “head.”

“ReDial” by Oregon (*Ecotopia*, 1987): The guitar riff (repeated later on the soprano saxophone) carries the typical “call and response” and “seemingly even, but uneven” characteristics of clave direction with the phrase on each side consisting of three plucked chords. Yet, the clave direction is *not* 3-3; it’s 2-3. It’s where the onsets fall that determines clave direction, rather than how many there are. (The clave direction is also hinted at, rather freely, by Trilok Gurtu’s interesting use of the *caxixi*.)

“Pa De Sekel” by Nestor Azerot (*MaxiCompas: Le Meilluer de la Musique D’Haiti*, 1998): vocals and keyboards in 3-2.

“Taxi Driver (live in Paris)” by Steel Pulse (reggae band from Birmingham, UK; from the album *Rastafari Centennial - live in Paris*, 1992): features a lengthy “toasting” section in 3-2 clave, with the clave outlined by the instruments in the background.

“Fabara” by Badenya—les frères Coulibaly (*Sénivè*, 2000, from Burkina Faso) starts with a driving African swing on the *dounun*. In the frantic pace, the 3-2 clave feel can be heard in various places in the song.

“Boroto” by Badenya—les frères Coulibaly (*Sénivè*, 2000, from Burkina Faso) is a crossover piece featuring *balafon*, *djembe*, talking drum, synthesizer, drum machine, and either electric or sequenced bass. The first half carries a greater sense of 3-2 clave while the second half—introduced by a scorching *djembe* call—is clave-neutral.

“Olufela” by Kayode Olajide (from Nigeria; off the compilation *SoundAffects: Africa*, CD 1, 2006): The slight change from one part of the intro drum beat to the next is a subtle statement of direction.

“Okuzua” by Orchestra Lissanga (from Zaire/DRC; from the compilation *Afro-Rock, Volume 1: A collection of rare and unreleased Afro-Beat quarried from across the continent*, 2002): in 3-2.

“Selma to Soweto” by Orlando Julius (from Nigeria; off the compilation *SoundAffects: Africa*, CD 1, 2006): in 3-2.

“El Baile del Buey Cansao” by Cuban *songo* pioneers Los Van Van (*Los Van Van*, 1969) is an example of a Cuban 2-3.

“A Cuba Volver” by Colombia’s Orquesta Guayacan (*A Puro Golpe*, 1994) puts the *son montuno* on a guitar (instead of piano) and accompanies it with a direct statement of the 2-3 *son* clave.

“Kaquiry Kaquiry” by Orquesta Guayacan (*A Puro Golpe*, 1994) reverts to *montuno* on the electric piano, and the underlying 2-3 clave is so clear, it is not explicitly stated.

“Ahmed Sabit” by Samba Mapangala & Orchestre Virunga (*Virunga Volcano*, 2008): electric guitar plays the “Brazilian” rhythm pattern of *samba de roda* and *samba-reggae*, (Figure 9, center, in Appendix A).

“Rock ‘N’ Roll Lovers” by Coryell/Mouzon (*Back Together Again*, 1977): 3:2 *son* on rhythm guitar.

“Olomage Ma Jo” by King Sunny Ade (*The Return of the Juju King*, 1990).

“Odesia” by Les Mangelepa (*Guitar Paradise of East Africa*, 1995): Sections in 2-3, with the percussion providing the clave-neutral *tresillo*, and the guitar indicating the same Afro-Brazilian pattern as Figure 9 (center) in Appendix A.

“Psychedelic Woman” by Honny And The Bees Band (Ghana): 3-2 guitar pattern (1101 | 1000 | 1010 | 1000).

“Bushi” by Gomez–Pellitteri–ESEN (*Trio*, 1986): 3-2 cymbal pattern, as well as the piano intro and much of the comping.

Glenn Miller's “In The Mood” has one of the most anti-*partido-alto* patterns possible (1011 | 1101 | 1011 | 1101).

H.3 Clave-type Timelines or Patterns found in Non-Clave Musics

“Faith” by George Michael (*Faith*, 1987): explicit 3-2 *son* clave on acoustic guitar and vocals.

“Superstition” by Stevie Wonder (*Talking Book*, 1972): the famous clavinet riff is mostly 2-3 “bossa” clave (rhythmically, not in terms of the pitches).

“How We Do (feat. 50 Cent)” by The Game (*The Documentary*, 2004): main-chorus vocals and synthesizer in 2-3 “bossa clave.”

“Message from Baghdad” by Acrassicauda (*Only The Dead See The End Of The War*, 2010): guitar riffs in 2-3 clave direction, including “bossa clave.”

“By The Time I Get To Arizona” by Public Enemy (*Apocalypse 91...The Enemy Strikes Black*, 1991): 3-2 “bossa clave” on vocals and turntables.

“Cadillac” by Bo Diddley (*Bo Diddley Is A Gunslinger*, 1960): 3-2.

“The Country of the Future” by Mirah from the album *(a)spera* features the author on *pandeiro*, *repique*, *surdo*, and *caixa*, along with Emily Kingan of The Haggard on *caixa* and Bryce Panic on *claves*. The vocals are clearly in 2-3, as accented by Bryce’s *clave*-playing. The *surdo* part is a *clave*-neutral *baião* approximation. The *pandeiro* part is inspired by *capoeira*. There are two *caixa* parts, one from Olodum-style *samba-reggae* (*clave*-neutral) and one based on the *Mangueira* samba school. Analysis of the latter is left as an exercise.

“Crash” by ohGr (*Undeveloped*, 2011; track 4 of misindexed CD): two-bar 2-3 pattern on guitar and kick drum (one bar in 2-3, followed by one neutral bar):
1010|1001|0010|1000, followed by 1010|1000|1010|1000

“Almost Again” by Strapping Young Lad from the album *The New Black* features snare accents that outline a rhythm pattern very similar to Brazilian *partido-alto* in its 2-3 version (more common in jazz than in samba).

“Two Minutes to Midnight” by Iron Maiden (*Powerslave*, 1984) is a typical example of the wide use of Brazilian 2-3 “bossa *clave*” in rock’n’roll, especially *NWOBHM* (New Wave of British Heavy Metal, which is no longer new, but is still known by that name). Quite a few *NWOBHM* songs feature this rhythmic pattern of power chords.

In contrast to the above, “Flash of the Blade” (also *Powerslave*, 1984) establishes *clave* direction in the drum beat, not on the guitars. This beat is also an example of 2-3, to the extent that it is used in the song (mainly in the intro).

“Still Life” by Iron Maiden (*Piece of Mind*, 1983): 2-3 “bossa *clave*” throughout the *vocals*!

“Paschendale” by Iron Maiden (*Dance of Death*, 2003): intro ride-cymbal pattern is 2-3 *samba de roda*.

“0-1” by Replikas (*avaž*, 2005): electric bass outlines the back half of 3-2 “bossa *clave*.”

“Sabah olsun” by seksendört (*seksendört*, 2005): guitar riffs are 2-3 “bossa *clave*.”

“Your Latest Trick” by Dire Straits (*Brothers In Arms*, 1985): cross-stick in 2-3:
0010|0100|0001|0010.

“We Disintegrate” by 16Volt (*Beating Dead Horses*, 2011): snare in 2-3, kick in 2-3, guitar in 2-3, and keyboard in 2-3.

“Keep ‘Em Straight (Intro)” by Balkan Beat Box (*Nu Med*, 2007): vocals in strict 2-3.

“99.9 F°” by Suzanne Vega (*99.9 F°*, 1992): chorus vocals in 2-3 *clave* direction.

“I Want Candy” by The Strangeloves (45-RPM single, 1965): 3-2.

“Bela Lugosi’s Dead” by Bauhaus (single, 1979): explicit 3-2 “bossa *clave*.”

“Canım kurban” by Erol Evgin (*İşte öyle bir şey*, 1977): 2-3 clave direction in the backing-vocal responses to the following the “canım kurban” call from the lead singer. At a higher level of rhythmic hierarchy, this song features a clave-like arrangement of accents in the primary refrain (with which the lead vocal enters), where measures of 7/4 and 4/4 alternate prior to the 4/4 call and response in clave. Furthermore, the backing-vocal response is (rhythmically) the same as Figure 9 (center) in Appendix A.

“Tavla” by Mirkelam (*Mirkelam*, 1995): main rhythm in 2:3 (based on the accent on the 16th preceding the downbeat of 3).

“Whoomp! There It Is” by Tag Team (*Whoomp! There It Is*, 1993): drum-machine bell part is 2-3 *samba de roda*.

“I’m Still In Love With You” by Al Green (*I’m Still In Love With You*, 1972): prominent hi-hat in 2-3.

“sanctified” by nine inch nails from the album *pretty hate machine* states the 3-2 *son* clave plainly on both the drum machine and the sequenced bass line.

“terrible lie” by nine inch nails (*pretty hate machine*, 1989): 2-3 for two of the synthesizer parts, clave-neutral elsewhere.

“Ambition” by Subway Sect (single, 1978): organ and some drum accents in 2-3.

“Break (techno mix)” by Turntable Terror (single, 1991): 2-3 *son* keyboard hits.

“You And I” by Queen (*A Day At The Races*, 1976): Piano *montuno* in 2-3.

“Stone Cold Crazy” by Queen (*Sheer Heart Attack*, 1974): Main riff in 2-3.

“Fat Bottomed Girls” by Queen (*Jazz*, 1977): Main riff in 2-3 (including all the variations on the album version not heard on the single version).

“Cool Cat” by Queen (Hot Space, 1982): Funk-style rhythm guitar in 2-3 *son* clave throughout. Although a number of sustained strums a 16th before the 1 mean that this song does not entirely follow clave direction, the strong presence of 2-3 *son* clave merits inclusion on this list. (Adhering to clave direction as a system of temporal harmony is not a requirement for any music that does not purport to be traditionally Afro-Latin.)

“More Of That Jazz” by Queen (*Jazz*, 1977) features a juxtaposition of 3-2 *son* on the drum set and one of the guitar riffs, and 2-3 “bossa clave” on the other guitar riff. This crossing of clave is fine because this slow rocker is not a samba or a rumba.

“We Will Rock You (Ruined by Rick Rubin)” by Queen (*BASIC beats sampler*, 1991): The turntable parts by Zulu DJ Afrika Islam and the kick drum by Chad Smith in the “ruined” section are all in 2-3 clave direction.

“The Invisible Man” by Queen (*The Miracle*, 1989): prominent synth-bass line through the song in 2-3 *partido-alto* (1110|1111|0111|1010).

“Alphabet Street (Indigo Nights/Live Sessions)” by Prince (*Indigo Nights/Live Sessions*, 2008) features guitar and bass in 2-3, and the main riff in 3-2 “bossa clave.” (Adhering to clave direction is not required for any music that does not purport to be Afro-Latin.)

“Hope For The Future” by Marillion (*This Strange Engine*, 1997): 2-3 hi-hat.

“Indianer” by Nena (*Nena*, 1983): prominent guitar-and-bass figure is 2-3 “bossa clave.”

“Zaubertrick” by Nena (*Nena*, 1983): “jam out” section three minutes into the song features prominent drum–guitar accents of 2-3 “bossa clave.”

“Noch Einmal” by Nena (*Nena*, 1983): guitar-and-keyboard riff in 2-3 (not “bossa”) featuring the isolated missing downbeat on 3.

“Keep Your Heart” by TV On The Radio (*Nine Types Of Light*, 2011): intro in 2-3, with the snare entering halfway through the 3-2 phrase, thus in clave with the 2-3 intro.

“West End Girls” by Pet Shop Boys (*DISCOGRAPHY: The complete singles collection*, 1991; original single 1985): Bell in 2-3.

“Love Comes Quickly” by Pet Shop Boys (*DISCOGRAPHY: The complete singles collection*, 1991; original single 1986): Main synth part in 2-3.

“Always On My Mind” by Pet Shop Boys (*DISCOGRAPHY: The complete singles collection*, 1991; original single 1987): Intro bell in 2-3.

“Ace of Spades” by motörhead (*Ace of Spades*): guitar licks and various vocal phrases are based on the 2-3 “bossa clave.”

“Girls On Film” by Duran Duran (*Duran Duran*, 1981): drum set in 3-2.

“Notorious” by Duran Duran (*Notorious*, 1986): vocals and guitar follow 3-2 *son* clave.

“colors of the fall” by Timescape (*two worlds*, 1997): “bossa clave” on *claves*.

“Be Gentle With Me” by The Boy Least Likely To (single, 2003): 2-3.

“My Mistakes” by Eleanor Friedberger (*Last Summer*, 2011): synth stabs in the same pattern as Oye Como Va’s famous organ riff in 2-3.

“On The 2nd Floor” by Airlock (*Drystar*, 2001): prominent synthetic ride cymbal in 2-3 clave throughout the song.

“The Color of Industry” by Radiation City (*The Hands That Take You*, 2011): 2-3 “bossa clave” on *claves*.

“Aeroplane” by Björk (*Debut*, 1993): 2-3 “bossa clave” on *claves*.

“Get Up (Before The Night Is Over)” by Technotronic (*Pump Up The Jam*, 1989): prominent synthesizer in 2-3.

“It Ain’t Over ‘Til It’s Over” by Lenny Kravitz (*Mama Said*, 1991): bass and guitar in 2-3.

“867-5309/Jenny” by Tommy Tutone (*Tommy Tutone-2*, 1981): vocals in 2-3 (1010|1110|1101|0110) on the chorus.

“World On Fire” by Krokus (*Change Of Address*, 1985): 2-3 “bossa clave” on bass riff during the intro, and on some of the subsequent rhythm-guitar parts.

“Fireball” by Deep Purple (*Fireball*, 1971): 2-3 “bossa clave” in the accents of the drum-set intro and some of the vocals.

“69” by Deep Purple (*Abandon*, 1998): Starts with 2-3 *partido-alto*.

“Stand Up And Shout” by Dio (*Holy Diver*, 1983): NWOBHM rhythm guitar in 2-3 “bossa clave.”

“747 (Strangers In The Night)” by Saxon (*Wheels Of Steel*, 1980): rhythm guitar in 2-3 *son*.

“Echo Chamber” by Beats International (*Echo Chamber*, 1991): 2-3 “bossa clave” on the synthesizer.

“Pahadi” by Alms For Shanti (*kashmakash*, 2007): 2-3 “bossa clave” in the beginning, 2:3 funk on snare later in the song.

“This Will Hurt Someone” by Dead World (*UHF/VHF*, 1995): drum machine in 2-3.

“Fly The Orient” by Tricky Woo (*A Fistful Of Rock N Roll, Vol. 4*): guitar and chorus vocals in 2-3 clave direction (with “isolated missing downbeats” around beat 3).

“Greathop [Dub]” by Mad Professor (*dubTronic*, 1998): 2-3 funk.

“Mona” by Quicksilver Messenger Service (*Happy Trails*, 1969): 3-2.

“Naked Rain” by This Picture (*a violent impression*, 1991): strings in 2-3 “bossa clave,” vocals and rhythm guitar in clave-neutral “*tresillo*” and related neutral patterns.

“Bright Stars” by Daedelus (*Rethinking The Weather*, 2003): whistling in 2-3 *partido-alto*.

“Heavensent” by Konkhra (*Weed Out The Weak*, 1997): guitar riff in 2-3.

“In The End” by Linkin Park (*[HYBRIDTHEORY]*, 2000): intro scratching in 2-3.

“Unbelievable” by EMF (*Schubert Dip*, 1990): guitar and piano (after each chorus) in 2:3.

“Advice” by Sly & The Family Stone (*A Whole New Thing*, 1967): 2-3 funk drum set.

“Down By The Water” by P J Harvey (*To Bring Yo My Love*, 1995): This piece features the drum set on the clave-neutral “tresillo” pattern throughout, but joined at some point by the 2-3 “bossa clave” pattern (played on the *claves*), establishing an overall sense of 2-3 clave direction. However, near the end of the song, with the line “little fish, big fish...” a new clave direction is (somewhat) established, where the vocal part suggests 3-2. (At this point, the 2-3 *claves* have stopped, so there is no mismatch.) The instrumental backing is indistinct, but suggests the same clave direction as the repeating outro vocal.

“Me And Julio Down By The Schoolyard” by Paul Simon (*Paul Simon*, 1972) and Simon & Garfunkel (*The Concert in Central Park*, 1981): main acoustic-guitar rhythm in 2-3.

H.4 Recommended Listening in African and Afro-Latin Music

- Ilê Aiyê – *25 Anos* (© 1999 Natasha CRCL-8812) Brazil.
- Monobloco – *2002* (©2002 MBcd1) Brazil.
- Monobloco – *Ao Vivo* (©2006 Som Livre 7891430021023) Brazil.
- Daniela Mercury – *Fejão Com Arroz* (© 1997 Sony Music Entertainment LLC (Brasil) I.C.L./Sony Latin CDZ 82118 2-479390) Brazil.
- Daniela Mercury – *Sol da Liberdade* (©2000 BMG Brasil Ariola 74321 73818-2) Brazil.
- Zeca Pagodinho – *acústico MTV* (©2003 Universal Music) Brazil.
- Michael Spiro-Mark Lamson – *Bata Ketu - A Musical Interplay of Cuba and Brazil featuring vocalists Bobi Cespedes & Jorge Alabé* (©Bembé Records 2011-2) USA, Cuba, Brazil.
- “Escola de Samba Mocidade Independente de Padre Miguel” – *Batucada* (©1990 JSL 003) Brazil.
- Badenya—les frères Coulibaly – Séniwè (©2000 Cod Music AG Trace 38801) Burkina Faso.
- Doudou NDiaye Rose – *Sabar* (© 2003 Mélodie Recordings) Senegal.
- Doudou NDiaye Rose – *Djabote* (©1992 RealWorld 7243 8 39470 2 7 Carol 2340-2) Senegal.
- Boudiabay NDiaye – *Gonyie Mouride Bamba* (Studio 2000, Dakar, Sénégal) Senegal.
- Doudou N’Diaye M’Bengue – *Thiato thiate* (© 2001 Felmay B004HS8BD4) Senegal.
- Le Ballet National Du Sénégal – *Pangols-The Spirit of West Africa in Music, Song, and Dance* (©1998 Le Ballet National Du Sénégal Musika 240898) Senegal.
- Deep Rumba/Rumba Profunda – *This Night Becomes A Rumba* (© 1998 AMERICAN CLAVE AMCL CD 1008/9) Cuba, USA, etc.
- Pancho Quinto – *En el Solar la Cueva del Humo* (©1997 Round-Wrold Music) Cuba.
- Los Muñequitos De Matanzas – *Rumba Cliente 88/77* (©1992 QBADISC QB-9005, Egrem, Artex) Cuba.
- Grupo Exploración – *drum jam* (©2000 Bembé Records 2026) USA, Cuba.

- Alex Acuña y su Acuarela do Tambores – *Rhythms for a New Millenium* (©2000 DCC Compact Classics, 185) Pan-American.
- Mongo Santamaria & friends – *mambo mongo* (©1993 Chesky Records D 101574) Cuba.
- Milton Cardona – *Cambucha (Carmen)* (© 1999 American Clavé Records AMCL 1028) Puerto Rico, Cuba.
- Giovanni Hidalgo – *Hands In Motion* (© 1997 Rmm Records, 82053) Puerto Rico.
- Giovanni Hidalgo & Horacio “El Negro” Hernandez – *Traveling Through Time* (DCI Music Video, 1996) Puerto Rico, Cuba.
- Giovanni & Changuito – *Conga Masters: The Masters Meet Again!* (DCI Music Video, 1996) Puerto Rico, Cuba.

APPENDIX I: Author's Musical Background and Qualifications

Since the author does not hold a degree in music, a summary of musical experience is included here as partial qualification of conducting research in music. As discussed in the opening of Appendix D, intimate engagement in music is considered a significant qualification for doing research in music.

The author has studied aspects of Brazilian, Cuban, Ghanaian, Middle Eastern and Japanese drumming with teachers Jorge Alabê (of Mocidade), Anderson Pandeiro (of Mangueira), Marcio Peeter & Wagner Profeta Santos (of Ilê Aiyê), Gamo Da Paz, Bruno Moraes & Alex Rangel (of Mocidade), Airto Moreira, Boca Rum, Curtis Pierre, Justino Roger, Emiliano Benevides and Jorge Martins [Brazil], Michael Spiro, Mark Lamson, and Scott Wardinsky [Cuba], Okaidja Afroso [Ghana], Michael Beach [Balkans/Middle East], and Torimaru Yumi and Kimura Kohei [Japan].

He is a founding member of the Japanese *dentou geinou* ensemble Takohachi, and a long-time member (and part-time co-director) of the acclaimed samba group Lions of Batucada, as well as long-time member of the Afro-Cuban-style dance company Axé Didé, and the rock band The Rotating Leslies. He is also the founder of the eclectic punk-jazz combo toyboat toyboat toyboat and the experimental samba group Mais Que Samba.

In 2005, an interpretation of his rhythm composition '*Baião Rumba a la Turc*' was included on the full-length CD release, *World of Percussion* by the Engin Gürkey Percussion Ensemble. Later, in 2009, the Mirah album (*aspera*) featured four layers of Vurkaç's Brazilian percussion on the song '*The Country Of The Future*' (also featuring members of The Decemberists and The Haggard), and rose to number 5 on the Rolling Stone college-radio chart.

Musicians Vurkaç has accompanied include Jorge Alabê, Pink Martini, Fishbone, Olodum, Mirah, Michael Spiro/Mark Lamson with Axé Didé, Airto Moreira, Tara Jane O'Neil and Obo Addy. In addition, he has shared concert bills (as part of various bands) with Smash Mouth, Storm & The Balls, My Name, Sage, Irma Thomas, Dirty Martini, Morcheeba, Sean Croghan, and Sean Lennon.

Vurkaç also taught the beginning-samba class for Lions of Batucada 2001–2009.

APPENDIX J: A Survey of Notational Conventions and Preferences for Meter and Time Signature for Musics of the African Diaspora

J.1 Purpose

The purpose of this survey is to establish—to an imperfect degree, since this is a subjective, cultural issue—preferred convention(s) for notating music of the African Diaspora in standard (universal) European music notation. The three aspects considered are the *tatum* (smallest subdivision [34]), time signature, and the number of measures assigned to one cycle of a clave pattern. Clearly, these three aspects are interconnected such that only a proper subset of all possible combinations is musically meaningful¹⁷⁷.

J.2 Data

The following sources were examined for this survey:

- a) *Ritmos do Brasil para Bateria* by Nenê¹⁷⁸ [1]
- b) *Brazilian Rhythms for Drumset* by Duduka Da Fonseca¹⁷⁹ & Bob Weiner¹⁸⁰ [2]
- c) *Aprendendo A Tocar O Bateria Carioca – As Baterias das Escolas de Samba do Rio De Janeiro* by Dr. Guilherme Gonçalves¹⁸¹ & Mestre Odilon Costa¹⁸² [3]
- d) *Brazilian Percussion Manual – Rhythms and Techniques with Application for the Drumset* by Daniel Sabanovich¹⁸³ [4]

¹⁷⁷ This study only concerns simple-metered rhythms. The corresponding question for compound-metered timelines in African and Afro-Latin music is only between 12/8 and 6/8. This is not discussed here.

¹⁷⁸ Professor of Brazilian Rhythms at the Universidade Livre de Música Tom Jobim (ULM), Nenê has played with and for Charlie Haden, Milton Nascimento, Steve Lacy, Paulo Moura, Kenny Wheeler, Hermeto Pascoal, Egberto Gismonti, Bob Moses, Pau Brasil, Marcos Suzano, and the “Pearl Masters” classes.

¹⁷⁹ Duduka Da Fonseca is the drummer of *Trio Da Paz*, one of Brazil’s leading authentic-Samba/Jazz combos.

¹⁸⁰ Bob Weiner has played with Harry Belafonte, Dianne Reeves, Hugh Masekela, and Herbie Mann.

¹⁸¹ Berklee alumnus and Professor of percussion at the Villa-Lobos Music School in Rio De Janeiro and the Brazilian Popular Music Conservatory of Curitiba, Dr. Gonçalves has performed with Brazil’s National Symphony Orchestra, the Symphony Orchestra of Parana, Rio Jazz Orchestra, Rio Dixieland Jazz Band, and published a method book entitled “*O Ritmo pelas Subdivisões*.”

¹⁸² Currently the director of the samba school G.R.E.S. Grande Rio and winner of two trophies for “Best Mestre” (1991, 1999), Odilon has played for Sergio Mendes, Caetano Veloso, Simone, Dionne Warwick, and top Rio samba schools G.R.E.S. Beija-Flor and G.R.E.S. Acadêmicos Do Salgueiro.

¹⁸³ Having written his master’s thesis on samba percussion, Sabanovich has played with Justo Almario, Charlie Byrd, Clare Fischer, Joe Henderson, Bobby Hutcherson, Woody Shaw, and studied with the Rio samba school, G.R.E.S. Beija-Flor.

- e) *Traditional Afro-Cuban Concepts in Contemporary Music* by Arturo Rodriguez¹⁸⁴ [5]
- f) *The Conga Drummer's Guidebook* by Michael Spiro¹⁸⁵ [6]
- g) *TimbaFunk by Talking Drums* by David Garibaldi¹⁸⁶, Jesús Diaz & Michael Spiro [7]
- h) *Songbook: Caetano Veloso, Volume 2*, produced and edited by Almir Chediak [8]
- i) *Spanish Montunos – Ensemble Studies in Rhythmic Precision and Pitch Security* by Ronald Herder [9]
- j) *Riddim: Claves of African Origin* by Billy Martin¹⁸⁷ [10]
- k) *A Mathematical Analysis of African, Brazilian, and Cuban Clave Rhythms* by Godfried Toussaint¹⁸⁸ [11]
- l) *Afro-Cuban Rhythms for Drumset* by Frank Malabe & Bob Weiner [12]
- m) *The Tomás Cruz Conga Method – Conga Technique as Taught in Cuba, Vols. 1, 2 & 3* by Tomás Cruz¹⁸⁹ [13]
- n) *Au des Cœur batteries de Rio – Méthode D'Initiation LE SAMBA de Enredo* by P. Nasse, Jean-Christoph Jacquin, B. Ginestet, and K. Blasquiz [14]
- o) *Evolution of the Tumbadoras* by José Luis Quintana (a.k.a. Changuito)¹⁹⁰ [15]
- p) *Conga Drumming & Afro-Caribbean Rhythms* by Jerry Gonzalez¹⁹¹ [16]
- q) *Inside the Brazilian Rhythm Section*, by Nelson Faria¹⁹² & Cliff Korman [17]
- r) *Brazilian Jazz Guitar* arranged by Mike Christiansen & John Zaradin [18]

¹⁸⁴ Arturo Rodriguez has played with Tito Puente and Dave Valentin, and studied with Senegalese sabar master Maphaté Diop, legendary Cuban conguero Miguel “Angá” Diaz, and famous Latin Jazz percussionist Luis Conté.

¹⁸⁵ Michael Spiro has recorded or performed with David Byrne, Cachao, Dori Caymmi, Changuito, Ella Fitzgerald, Gilberto Gil, Giovanni Hidalgo, Toninho Horta, Bobby Hutcherson, Dr. John, Bobby McFerrin, Andy Narell, Chico O'Farrill, Eddie Palmieri, Lázaro Ros, Carlos Santana, Omar Sosa, Clark Terry, McCoy Tyner and Charlie Watts.

¹⁸⁶ David Garibaldi has recorded or performed with Patti Austin, Natalie Cole, Larry Carlton, Ray Obiedo, the Buddy Rich Orchestra, Boz Scaggs, Deniece Williams, Tower of Power, and the Yellowjackets.

¹⁸⁷ Billy Martin of Medeski, Martin & Wood has also played with John Scofield, John Zorn, Iggy Pop, and Chuck Mangione.

¹⁸⁸ According to his autobiography, Toussaint has “been a professor in the School of Computer Science at McGill University since 1972 [and his] primary research interests are in the design and analysis of algorithms for [...] pattern recognition, music information retrieval, computational music theory, [etc.].” (<http://cgm.cs.mcgill.ca/~godfried/>). Accessed 1/11/2007.

¹⁸⁹ Tomás Cruz has played with the Cuban ensembles of Pualito FG y su Élite, Juan Cerito, and Manolín.

¹⁹⁰ One of the most influential Cuban percussionists ever, Changuito is co-inventor of the *songo* style made popular by *Los Van Van*.

¹⁹¹ Jerry Gonzalez has played with Dizzy Gillespie, Eddie Palmieri, Steve Berrios, Tony Williams, McCoy Tyner, Anthony Braxton, Ray Barretto, Tito Puente, Paquito D'Rivera, and Machito.

¹⁹² Faria, a Brazilian national, has been a long-term accompanist for the legendary João Bosco.

- s) *The Brazilian Guitar Book – Samba, Bossa Nova and Other Brazilian Styles* by Nelson Faria [19]
- t) *Salsa Guidebook for Piano and Ensemble* by Rebeca Mauleón [20]
- u) *Salsa Hanon: 50 Essential Exercises for Latin Piano* by Peter Deneff [21]
- v) *The Latin Real Book: The Best Contemporary & Classic Salsa, Brazilian Music, Latin Jazz (C Version)*, edited by Chuck Sher, Larry Dunlap and Rebeca Mauleón-Santana [22]
- w) *101 Montunos* by Rebeca Mauleón-Santana [23]
- x) *El toque de Candombe – Ritmo tradicional Afro-Uruguayo* by Hugo “Foca” Machado, Willy Muñoz & Jorge Sadi [24]
- y) *Drum Gabu: The Rhythms of West African Drumming* by David Locke [25]
- z) *The Bongo Book* by Trevor Salloum¹⁹³ [26]
- aa) *How To Play Repinique* by Kurt Rasmussen¹⁹⁴ [35]
- bb) *African Musical Symbolism in Contemporary Perspective* by John Collins¹⁹⁵ [38]
- cc) *Pandeiro Brasileiro* by Louis Roberto Sampaio¹⁹⁶ & Victor Camargo Pub¹⁹⁷ [39]
- dd) *Rhythms and Songs from Guinea* by Famoudou Konaté¹⁹⁸ & Thomas Ott [40]
- ee) *New Ways Of Brazilian Drumming – Nuevos Caminos de la Batería Brasileña* by Sergio Gomes¹⁹⁹ [41]
- ff) *The ABCs of Brazilian Percussion: The Easiest Way to Teach Yourself How to Play the Essential Brazilian Percussion Instruments* by Ney Rosauero²⁰⁰ [42]
- gg) *The 3:2 Relationship as the Foundation of Timelines in West African Musics*²⁰¹ by Eugene Domenic Novotney [43]

¹⁹³ Salloum has studied with some of the greatest *songo*, *rumba* and jazz musicians from Cuba, such as Irakere, Roberto Vizcaino, Memo Acevedo, and the rumba legends Los Muñequitos de Matanzas.

¹⁹⁴ Rasmussen is an official educator for the Latin Percussion (LP) company. He has studied with Carlinhos Pandeiro de Ouro, Mark Lamson, Marcos Suzano, Carlinhos Brown and members of such escolas as Vai-Vai and Mocidade Independente de Padre Miguel.

¹⁹⁵ Performing in Ghana since 1969, Collins has earned a Ph.D. in Ethnomusicology at SUNY Buffalo and became the head of the Music Department at the University of Ghana in 2003. He is currently a consultant for Ghanaian music unions and a performer of *Highlife* music.

¹⁹⁶ Brazilian composer and performer Luiz Roberto Cioce Sampaio has been professor of percussion at several universities throughout the northern regions of Brazil.

¹⁹⁷ Brazilian symphonic percussionist and drum-set artist Victor Daniel Camargo Bub has studied with Marcos Suzano and Marcio Bahia and has been a member of numerous orchestral and experimental percussion groups throughout Brazil.

¹⁹⁸ Guinean drummer Konaté has been a member of both *Ballet Africains* (a multinational group) and the *Ballet Africains de la République de Guinée* since the 1960's.

¹⁹⁹ Sergio Gomes is a classically trained percussionist from São Paulo State University and a percussion and theory teacher at two Brazilian colleges. His Jazz ensemble, *Terra Brasil*, was nominated for a Grammy in 2004.

²⁰⁰ Rosauero has performed with Evelyn Glennie and the London Symphony, holds a doctorate degree in music, and is the Director of Percussion Studies at the University of Miami in Coral Gables, FL.

²⁰¹ This is Novotney's Ph. D. dissertation for the degree of Doctor of Musical Arts.

J.3 Findings

In the United States, perhaps due to the influence of jazz, or perhaps due to ease of sight reading, clave is commonly notated as a two-bar phrase, with the smallest subdivision being the 8th note. In studying Brazilian, Cuban, Ghanaian, and other Latin American and West African music, the author has found that (whether notation is employed or not), the tactus (beat) is generally thought of as being divided into four subdivisions²⁰². There are four strikes on the *tamborim* for each beat of the *surdo*; there are four parts to the ‘heel-toe’ motion of the basic *tumbao* patterns of salsa and *songo*. This leads many performers and teachers, especially of Afro-Brazilian music, to express the rhythm in terms of 16th-note subdivisions (discussed shortly).

Not surprisingly, it was found that Rodriguez, Nenê, Sabanovich, Fonseca, Locke, and the team of Dr. Gonçalves and Mestre Odilon all prefer 16th-note subdivisions. Dr. Toussaint and Billy Martin show the clave form as a set of 16 onsets (and rests), each primarily using their own form of IOI²⁰³-based notation, without alluding to any note values.

On the other hand, while Rodriguez, in transcribing Cuban music, uses one bar of 4/4 time with the TUBS notation, five of the six Brazilian-music books prefer 2/4 time, 16th-note tatum, and 2 bars. Brazilian percussionists Sampaio and Bub notate samba, *choro*, *frevo*, and *partido-alto* in 2/4 with 16th-note subdivisions, and *maracatú* in 4/4 with 16th-note subdivisions. (It is important to note that, as elaborated below, the difference between two bars of 2/2 and one bar of 4/4 is not as meaningful in African-derived music as it is in European music where the time signature reflects the accent structure to a much greater extent. Hence, for African and Afro-Latin music, two bars of 2/2 with a 16th-note tatum is the same as one bar of 4/4 with a 16th-note tatum.)

This interpretation is supported by Locke’s assertion that:

Gahu seems to be in 4/4 time, but this time signature implies inappropriate conventions of accentuation—strongest accent to the first stress, secondary accent to the third stress, and weak accent to the second and [fourth] stresses. It might be closer to an African perception to regard the beat as a tactus, that is, an unaccented organizational device, for in fact each beat receives an equal accent. [25, p. 19]

²⁰² Only simple (duple) meters are considered here. A survey of compound meters, involving 6/8 or 12/8 bell patterns, are left for another study.

²⁰³ Interonset Interval, the musical distance between the attacks of successive notes.

Locke justifies his choice of one bar for the gankogui phrase (the original, Ghanaian equivalent of clave) by the following: “the key to timing in this music lies in the relationship of strokes in the gankogui phrases to all other actions” [25, p. 9]. Thus, Locke places the bar lines so as to have each *gankogui* (Ghanaian bell) phrase (clave) constitute one bar.

Salloum explicitly states that “in Cuba, clave is usually written as a one-bar phrase; however, in North America most musicians write clave as a two-bar phrase” [26, p. 8]. He goes on to write each cycle of clave in one bar of 4/4, with 16th-note subdivisions. One-bar cycles of 4/4 (with the smallest subdivision per pattern being 16th notes) appear throughout his book, although both representations are encountered.

The remaining Afro-Cuban books (those by Spiro, Garibaldi and Herder), all feature two bars of 4/4 time with 8th-note subdivisions, as seen in standard North American jazz transcriptions.

Authors of the instructional book on Uruguayan *Candombe* (Yoruba religion and ceremonial music as manifested in Uruguay, not the same as *Candomblé* in Brazil), Machado, Muñoz and Sadi, describe the interchangeability of 2/2 and 4/4 time for Afro-Cuban, Afro-Brazilian, and Afro-Uruguayan clave (*madera*):

The *clave* of the traditional *Cuban Son* is transcribed in two measures of 2/4 time. Nowadays, the notation of Afro-Cuban rhythms is usually written in either 2/2 or 4/4 time [24].

The examples they use for clave alone are based on 8th-note subdivisions [24, p. 10], while the *madera* is shown as one bar of 4/4 time with a 16th-note subdivision [24, p. 13]. All subsequent music notation in the book (for the Afro-Uruguayan style of *Candombe*) is given with 16th-note subdivisions.

A number of other instructional books also consistently imply that while both notational approaches are acceptable, the 16th-note approach is either equally acceptable or more accurate. For example, *The Tomás Cruz Conga Method – Conga Technique as Taught in Cuba, Vols. 1, 2 & 3* use 16th-note-based notation almost exclusively, and state in no uncertain terms that “in Cuba, music is usually written in 16th notes such that one clave lasts one measure of 4/4” [13, Vol. 1, p. 63], and that “in terms of describing what’s really taking place in the music there’s no doubt that the 16th note approach is correct” [13, Vol. 2, p. 18].

Furthermore, Cruz, et al. go on to explain in the clave appendix to their second volume that:

the 16th note method really makes more sense because it shows the rhythm over the space of 4 quarter notes, and, in common time, or 4/4, a quarter note gets one beat. If you were to take the 8th note diagram literally, you would conclude that one clave lasts 8 beats, each with 2 subdivisions, which [...] is completely wrong and will result in a beginner feeling the clave incorrectly. [13, Vol. 2, p. 78]

Similarly, on Latin Pulse Music's timba.com, one of Cruz's co-authors, Kevin Moore presents both sides, but settles on 16th notes:

Most Cubans write it in 16ths and most others write it in 8ths. The 16th note camp argues that their method shows the pulse where it really is. The 8th note camp insists that 8ths are easier to read on the gig. In defense of the 8th note devotees, if you play it with the correct feel, it doesn't matter how you write it, and experienced musicians can read 8th notes while feeling the pulse in half notes. We'll go along with them on this, and have no problem with writing parts for experienced musicians in 8th notes, but to learn to *understand* clave in the first place, it's horribly misleading to display it as 8th notes spanning two bars because it draws attention away from the real pulse. Our solution will be to show it graphically [a combination of TUBS and X-O notation shown below this text in the source], which combines ease of reading with the 16th note conceptualization. [27]

Likewise, Sabanovich emphasizes the 16th-note subdivision repeatedly: "there will be a constant 16th-note flow" [4, p. 7], "keep the accented 16th notes very strong" [4, p. 8], and "always emphasize the 4th and 5th 16th notes with an accent." [4, p. 11]

Afro-Cuban Rhythms for Drumset by Frank Malabe and Bob Weiner—while mainly written with a preference for 8th-note increments—states explicitly (in the section called "clave in 4/4 time") that the notation of clave with 8th-note and 16th-note subdivisions are completely interchangeable: Although you would usually see this clave written in 8th notes, you could also write it in 16th notes. [12, p. 19]

Similarly, Kurt Rasmussen, an educator for Latin Percussion™, prefaces his use of two bars of 2/4 by stating that a single bar of 4/4 (with 16th-note subdivisions) is

most common, but that he is specifically emphasizing the antecedent/consequent nature of clave by his choice of two bars [35].

Note that seven of the nine specifically Brazilian books and four (and partially another) of the seven Cuban sources used 16th notes with 2/4 time. Among the books of sheet music available for this study, one covered three possibilities using both 8th- and 16th-note subdivisions as the notated subdivisions; another notated Brazilian music for guitar using 16th-note subdivisions with 4/4 meter; and the remaining two (by Brazilian guitarist Nelson Faria) feature 2/4 meter with 16th-note subdivisions. Of those, *The Brazilian Guitar Book* [19] displays songs by Jobim, Milton Nascimento, Moacyr Santos, Baden Powell, and João Bosco in 2/4 time with consistent 16th-note subdivisions.

Consequently, neither the prominence of 8th-note subdivisions in music notation for jazz, bossa nova, and salsa, nor the predominance of 16th-note written subdivisions for the folkloric forms can be denied. The reason for this is most likely in the nature of folkloric percussion in Cuba and Brazil, and the connections that salsa and bossa nova enjoy with the jazz music of North America.

In the Cuban case, as shown in [15, 16], the smallest standard subdivision of time is the heel/toe (or, palm/tip), where the right palm, left palm, right fingertips, and left fingertips, respectively, sound the drum for each tatum. This is true for all styles utilizing some variation of the heel-toe *tumbao*, such as *son*, *salsa*, *songo* and *rumba*.

Likewise, in folkloric Brazilian music, the smallest essential subdivision is the *repinicado* rhythm (rolling samba 16ths) played on the *tamborim*, *pandeiro*, *chocalho*, and *repique*. On each instrument, this is a pattern of four strokes. For the *tamborim*, it consists of the downbeat ('1'), forward-strike grace note ('e'), flip grace note ('&'), and the driving stroke, or pick-up ('a'). On the *pandeiro*, one common way to play this pattern is thumb ('1'), tip ('e'), palm ('&'), slap ('a'). Similarly, on the *repique*, the first stroke is a stick tone, the second and third constitute a controlled rim-shot bounce, and the fourth is a hand slap. A related sequence of four movements apply to the shaking of the *chocalho*. In all these cases, the four strokes take place in the time of one tactus (beat) by the first or second *surdo* players, who provide the basic pulse of samba. Thus, the basic pulse is divided into four critical and ubiquitous subdivisions.

Gomes addresses the difference between common ('northern' or jazz) notation and the proper way of thinking about samba in justifying his use of the less common, but more Brazilian notational system: "To practice the phrases, think of systems in 4/4, because they describe quaternary cycles, even though samba is usually written in 2/4" [41, p. 23].

What follows is a table summarizing the notational choices made by the educators, researchers, and authors whose work was surveyed. The numbers shown in the table entries are the numbers of books and scholarly articles using each notational convention. The fractional numbers are for those cases in which the author or editor uses more than one convention. (It should be noted for printing purposes that the table is color-coded, as indicated in the caption.)

Even though this admittedly simple analysis is based on a convenience sample, that sample is nonetheless quite large (over 30 sources in the form of books, articles, online lessons, and instructional video recordings, including several written by professional Brazilian musicians and Brazilian music scholars) for a topic that only recently has been receiving attention in the music-learning literature.

It is seen that 70% of the sources used 16th-note subdivisions, 27% used 8th-note subdivisions, and 3% did not specify a note value. The general preference is for 16th notes.

As for the choice of time signature, 55% used 4/4, 36% used 2/4, 3% used 2/2, and 3% used 16/8. (The remaining 3% did not specify a time signature.) The overall preference is for 4/4 time (when ignoring compound meters).

Finally, for the number of measures used to represent one clave cycle, 61% chose two bars, and 39% chose one bar, including the source that did not specify time signature, but displayed clave phrases in one piece. The preference is for two bars, but with a smaller margin than in the case of subdivisions.

The best choice appears to be two bars of 4/4 time with 16th-note subdivisions. Table 21 shows the breakdown of standard-notation preferences for clave rhythms: Red entries are for two-bar representations; blue entries are for one-bar representations. Shown here are the main choice each author makes for notation.

Table 23: Choice of Subdivisions

	2/4	4/4	2/2	16/8	none
8 th notes		7	1	1	
16 th notes	12	11			
unspecified					1

A slightly more involved analysis can be undertaken if all the representations encountered in each source are included (Table 22). In this case, a breakdown of both the primary form of notation and the alternatives presented by each author is included. In the case of authors who use more than one form, the different forms are listed with equal weighting.

Table 24: Choice of Subdivisions

	2/4	4/4	2/2	16/8	none
8 th notes		7+ $\frac{1}{3}$	$\frac{2}{3}$	1	
16 th notes	11+ $\frac{1}{3}$	11+ $\frac{1}{3}$	$\frac{1}{3}$		
unspecified					1

Once again, it is seen that 70% of the sources used 16th-note subdivisions, 27% used 8th-note subdivisions, and 3% did not specify a note value. The general preference is for 16th notes.

As for the choice of time signature, this time we find that 57% used 4/4, 34% used 2/4, 3% used 2/2, and 3% used 16/8. The general preference is for 4/4 time.

Finally, for the number of measures used to represent one clave cycle, 59% chose two bars, and 41% chose one bar, including the source that did not specify time signature, but displayed clave phrases in one piece. The preference is for two bars.

Once again, the most popular combination is two bars of 4/4 with 16th-note subdivisions. However, this combination is not one that actually shows up in any of the sources surveyed, and, having 32 onsets per clave cycle, would indeed lead to a rather strange way of notating clave-based rhythms. This demonstrates that a direct analysis of these proportions of usage can at best be a general indication of the state of affairs.

It would, however, be poignant to either 1) preserve the unity of the clave phrase by writing it within one bar, and preserve the feel of the music by writing it with 16th-note subdivisions, or 2) to preserve ease of sight reading and emphasize the antecedent/consequent-nature of clave by using 8th notes and two bars of 4/4 time.

The complete list of notational methods, tools, and conventions encountered thus far by the author are enumerated below using a pattern common to Afro-Brazilian music, sometimes called bossa²⁰⁴ clave or Brazilian clave [41, p. 90; 42, p. 7].

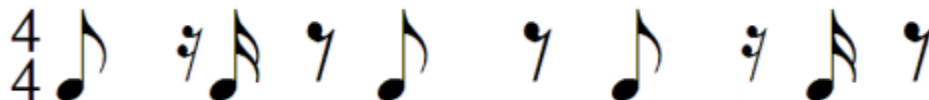
- 1) Standard (European) Notation with 2 bars of 4/4 time, using an 8th-note subdivision



- 2) Standard (European) Notation with 2 bars of 2/4 time, using a 16th-note subdivision

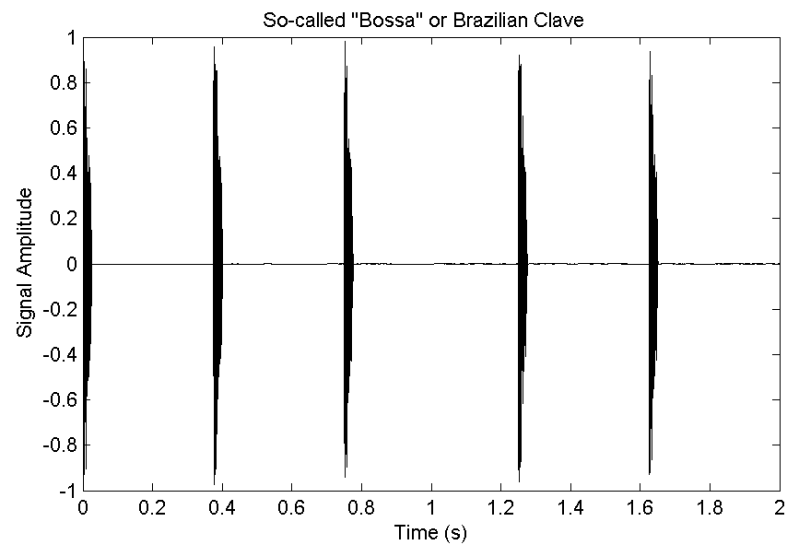


- 3) Standard (European) Notation with 1 bar of 4/4 time, using a 16th-note subdivision



²⁰⁴ The terms “bossa clave” and “Brazilian clave” are misnomers commonly used by North Americans who study Brazilian music. The term “clave” is a Spanish word, and its musical use in terms of rhythm is of Cuban origin. Since the concept underlying clave is African, it is somewhat inaccurate to refer to African and non-Cuban Afro-Latin music with the term clave. Many other terms and descriptions are also in use, though none are as commonly recognized or as concise: *okele* [37], *madura* [24], *clips* [38, p. 31], *compás* [31], tension and release (a very common description), the principle of mobility and finality, etc. As music terms go, “bossa clave” is akin to “English horn,” which is neither English nor a horn [36]. The “Brazilian clave” is neither solely Brazilian (it is also African), nor exactly a clave pattern, though the very subtle differences between Brazilian and Cuban clave are beyond the scope of this work.

4) Audio waveform display



5) TUBS, with the count from notation (3) above it, and notation (2) under it

1	e	&	a	2	e	&	a	3	e	&	a	4	e	&	a
X			X			X				X			X		
<u>1</u>	&	2	&	3	&	4	&	<u>1</u>	&	2	&	3	&	4	&

6) Circular TUBS [38, pp. 58, 59] and acoustic mandalas [Ibid., pp. 66–68] (not shown here, but these are graphic combinations of forms 5, 7 and 12)

7) Binary Notation (related to TUBS)

1001001000100100

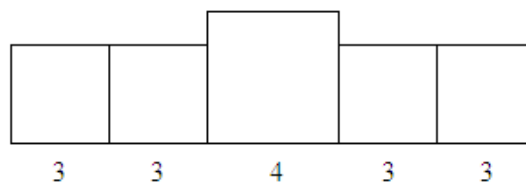
8) X-dot/X-O Notation (related to TUBS)

[X . . X . . X . . X . . X . .]

9) IOI sequence

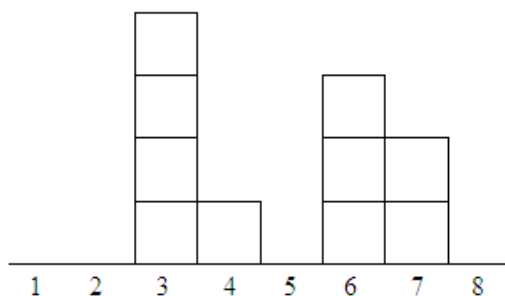
3 – 3 – 4 – 3 – 3

10) TEDAS²⁰⁵ (“olympic²⁰⁶”) representation (related to IOI) [11]

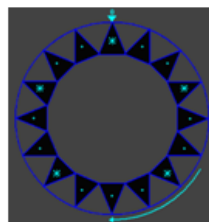
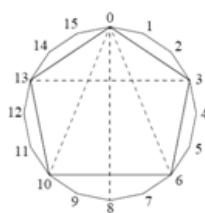


11) Interval Vector and Interval-Spectrum Histogram (related to IOI) [11, 29, 30]

0 – 0 – 4 – 1 – 0 – 3 – 2 – 0



12) Convex Polygons [11, 30, 31] or Cyclic Arrays [33], left; and Rhythm Wheels [32], right.



²⁰⁵ Temporal Elements Displayed As Squares [28].

²⁰⁶ So called by the author because the sizes of the squares represent the inter-onset interval to the next note attack, and, as a result, the TEDAS representation resembles the medal podium at the Olympic Games.

Many of these representations have been discussed in detail by Toussaint [11, 30, 31] and others. Our purpose here is, rather, to compare the frequencies of usage and the musical advantages conferred by the notational alternatives.

J.4 Conclusion

For musicological, music educational and Music Information Retrieval (MIR) purposes, the methods recommended in this study are:

- 1) Standard (European) Notation with 1 bar of 4/4 time, using a 16th-note subdivision
- 2) Standard (European) Notation with 2 bars of 4/4 time, using an 8th-note subdivision
- 3) TUBS
- 4) Binary Notation (related to TUBS)
- 5) IOI Sequence
- 6) TEDAS

It is possible that the sample used for this survey consists of a set of resources among which expressing Afro-Latin rhythm patterns using 16th-note subdivisions turned out to be more common. This should not *necessarily* lead to the conclusion that the same level of preference would be found if *all* instructional and sheet-music sources were to be consulted. However, it does suggest that 8th-note and 16th-note increments, 1-bar and 2-bar representations, and a variety of corresponding time signatures may be acceptable. These representations are indeed interchangeable because of the reduced meaning (in terms of accent structure) of time signatures and note values when imposing European notation on non-European music. Hence, regardless of the use of 1-bar or 2-bar representations and an 8th-note or 16th-note written subdivision, all of the 16 events (presence or absence of onsets) that make up the clave cycle²⁰⁷ must be taken into account at once in computational analysis of clave.

It is concluded here that using multiple representations and basing technical development on a note-value-independent depiction common in MIR literature is the most effective tradeoff. This allows one to avoid compromising Euro-American musical convention (2 bars) and the conceptual core of the proposal's argument that clave signifies relative degrees of tension and resolution, rather than each half signifying an absolute rhythmic characteristic.

²⁰⁷ This only takes into account duple-time clave, not compound-meter Afro-American patterns.

References for Appendix J

- [1] Nenê, *Ritmos do Brasil para Bateria*, São Paulo, Brazil: Trama Editorial, 1999.
- [2] Da Fonseca, D., Weiner, B., and Riley, J., *Brazilian Rhythms for Drumset*, Van Nuys, CA: Alfred Publishing Co, 1991.
- [3] Gonçalves, G., and Costa, O., *Aprendendo A Tocar O Batuque Carioca – As Baterias das Escolas de Samba do Rio De Janeiro*, Brazil: Groove, 2000.
- [4] Sabanovich, D., ed. Cirone, A. J., *Brazilian Percussion Manual – Rhythms and Techniques with Application for the Drumset*, 2d. ed., Van Nuys, CA: Alfred Publishing Co, 1996.
- [5] Rodriguez, A., *Traditional Afro-Cuban Concepts in Contemporary Music*, Pacific, MO: Mel Bay Publications, 2003.
- [6] Spiro, M., and Ryan, J., *The Conga Drummer's Guidebook*, Petaluma, CA: Sher Music, 2006.
- [7] Garibaldi, D., Diaz, J., and Spiro, M., *TimbaFunk by Talking Drums*, Miami, FL: Warner Bros. Publications, 1999.
- [8] Chediak, A. (prod./ed.), *Songbook: Caetano Veloso, Volume 2*, 5th ed., Petrópolis, RJ, Brazil: Lumiar Editora, 1994.
- [9] Herder, R., *Spanish Montunos – Ensemble Studies in Rhythmic Precision and Pitch Security – Solo Book of Montuno Improvisation for Concert Band, Stage Band, Orchestra or any Instrumental Combination*, New York, NY: Continuo Music Press, 1972.
- [10] Martin, B., *Riddim: Claves of African Origin*, Columbus OH: Music In Motion Films Ltd, 2006.
- [11] Toussaint, G. T., *A Mathematical Analysis of African, Brazilian, and Cuban Clave Rhythms*, Montréal, Québec, Canada: McGill University, 2002.
- [12] Malabe, F., and Weiner, B., *Afro-Cuban Rhythms for Drumset*, Manhattan Music, 1990.
- [13] Cruz, T., Moore, K., Gerald, M., and Fiol, O., *The Tomás Cruz Conga Method – Conga Technique as Taught in Cuba, Vols. 1, 2 & 3*, Pacific, MO: Mel Bay, 2004.
- [14] Nasse, P., Jacquin, J.-C., Ginestet, B., and Blasquiz, K., *Au des Cœur batteries de Rio(in the heart of Rio's samba school baterias) – Méthode D'Initiation LE SAMBA de Enredo Percussions du Brésil*, France: Le Salon De Musique, 2002.
- [15] Quintana, J. L., *Evolution of the Tumbadoras*, Miami, FL: Warner Bros. Publications, 1996.

- [16] Gonzalez, J., et al., *Jerry Gonzalez & Fort Apache Band: Conga Drumming & Afro-Caribbean Rhythms*, Alchemy Pictures, 1994.
- [17] Faria, N., and Korman, C., *Inside the Brazilian Rhythm Section*, Petaluma, CA: Sher Music, 2001.
- [18] Christiansen, M., and Zaradin, J. (arr.), *Brazilian Jazz Guitar*, Pacific, MO: Mel Bay and Warner Bros. Publications, 2004.
- [19] Faria, N., *The Brazilian Guitar Book – Samba, Bossa Nova and Other Brazilian Styles*, Petaluma, CA: Sher Music, 2005.
- [20] Mauleón, R., *Salsa Guidebook for Piano and Ensemble*, Petaluma, CA: Sher Music, 1993.
- [21] Deneff, P., *Salsa Hanon: 50 Essential Exercises for Latin Piano*, Milwaukie, WI: Musicians Institute Press & Hal Leonard, 1997.
- [22] Sher, C., Dunlap, L., and Mauleón-Santana, R., eds., *The Latin Real Book: The Best Contemporary & Classic Salsa, Brazilian Music, Latin Jazz (C Version)*, Petaluma, CA: Sher Music & Hal Leonard, 1997.
- [23] R. Mauleón-Santana, *101 Montunos*, Petaluma, CA: Sher Music Co., 1999.
- [24] Machado, H. “Foca,” Muñoz, W., and Sadi, J., *El toque de Candombe – Ritmo tradicional Afro-Uruguayo / Traditional Afro-Uruguayan Rhythm*, Pacific, MO: Mel Bay Publications, Inc., 2002.
- [25] Locke, D. *Drum Gabu – The Rhythms of West African Drumming*, Crown Point, IN: White Cliffs Media Company, 1987.
- [26] Salloum, T. *The Bongo Book*, Pacific, MO: Mel Bay Publications, Inc., 1997.
- [27] http://www.timba.com/fans/clave_debates.asp
Accessed January 12, 2007.
- [28] Gustafson, K., “The Graphical Representation of Rhythm,” *Progress Reports from Oxford Phonetics*, Vol. 3, 1988, pp. 6–26.
- [29] Johnson, T. A., *Foundations of Diatonic Theory: A Mathematically Based Approach to Music Fundamentals*, Emeryville, CA: Key College Pub., 2003.
- [30] Toussaint, G., “The Geometry of Musical Rhythm,” *Discrete and Computational Geometry, Japanese Conference, JCDCG 2004, Revised Selected Papers*, Tokyo, Japan: Springer, 2004, pp. 201–204.
- [31] Díaz-Báñez, J. M., Farigu, G., Gómez, F., Rappaport, D., and Toussaint, G. T., “El Compás Flamenco: A Phylogenetic Analysis,” *Proceedings of BRIDGES*:

Mathematical Connections in Art, Music and Science, Winfield: Kansas: Southwestern College, 2004, pp. 61–70.

- [32] http://www.ccd.rpi.edu/Eglash/csdt/latino/rhythm/rhy_intr.htm
- [33] McLachlan, N., “A Spatial Theory of Rhythmic Resolution,” *Leonardo Music Journal*, Vol. 10, 2000, pp. 61–67.
- [34] Seppänen, J., “Tatum grid analysis of musical signals,” *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 2001, pp. 131–134.
- [35] Rasmussen, K., “How To Play Repinique,” *Latin Percussion: Online Video Lessons*, Available: www.lpmusic.com/Play_Like_A_Pro/Tips/kurt.html Accessed August 16, 2007.
- [36] Kerman, J. and Kerman, V., *Listen – Brief Edition*, New York, NY: Worth Publishers, Inc., 1987, p. 44.
- [37] Ekwueme, L. E. N., “Concepts of African Musical Theory,” *Journal of Black Studies*, Vol. 5, No. 1, 1974, p. 45.
- [38] Collins, J., *African Musical Symbolism in Contemporary Perspective*, Berlin: Pro BUSINESS GmbH, 2004.
- [39] Sampaio, L. R. and Bub, V. C., *Pandeiro Brasileiro*, Florianópolis, Brazil: Bernúncia Editora, 2006.
- [40] Famoudou, K., and Ott, T., *Rhythms and Songs from Guinea*, Oldershausen, Germany: Institut für Didaktik populärer Musik, 1997.
- [41] Gomes, S., *New Ways Of Brazilian Drumming – Nuevos Caminos de la Batería Brasileña*, Tübingen, Germany: advance music, 2007.
- [42] Rosauero, N., *The ABCs of Brazilian Percussion: The Easiest Way to Teach Yourself How to Play the Essential Brazilian Percussion Instruments*, Brazil: Propercussão Brasil, 2004.
- [43] Novotney, E. D., *The 3:2 Relationship as the Foundation of Timelines in West African Musics*, Urbana, IL: Graduate College of the University of Illinois at Urbana-Champaign, 1998.

APPENDIX K: The *automate.dat* Script for Neural-Net Experiments in NeuralWare's NeuralWorks

```
open ABE-b.nnd
initialize
randomseed 85
set learnfile Train.txt
learn 50000
set testfile Holdout.txt
test
savenet Holdout_ABE-b_85.nnd
```

APPENDIX L: The *AutomateNeuralWorks.vbs* Script for Exercising Network Configurations with Multiple Seeds

This script was written by Eric Egalite and customized by Mehmet Vurkaç.

```
Const logFilename="neuralWorks.log"

Set fileSystem = CreateObject("Scripting.FileSystemObject")

Dim mostRecentSeed
Dim mostRecentNeuralNetFilenamePrefix

'-----
' Returns an array containing each line read from the specified file
'-----
Function GetListFromTextFile(filename)

    Set textFile = fileSystem.OpenTextFile(filename)

    textFileContents = textFile.ReadAll

    textFile.Close

    GetListFromTextFile = Split(textFileContents, vbCrLf)

End Function

'-----
' Returns an array containing the filename of each neural net
' specified in neuralNetFilenameList.txt
'-----
Function GetNeuralNetFilenameList

    GetNeuralNetFilenameList = GetListFromTextFile("neuralNetFilenameList.txt")

End Function

'-----
' Returns an array containing each seed value specified in
' seeds.txt
'-----
Function GetSeedList

    GetSeedList = GetListFromTextFile("seeds.txt")
```

End Function

```
'-----  
' Generates a NeuralWorks automation script named automate.dat.  
'-----  
Sub GenerateNeuralWorksAutomationScript(neuralNetFilename, seed)  
  
    Set textFile = fileSystem.CreateTextFile("automate.dat")  
  
    Dim neuralNetFilenamePrefix  
    neuralNetFilenamePrefix = Replace(neuralNetFilename, ".nnd", "")  
  
    textFile.WriteLine("open " + neuralNetFilename)  
    textFile.WriteLine("initialize")  
    textFile.WriteLine("randomseed " + seed)  
    textFile.WriteLine("set learnfile 'Train.txt")  
    textFile.WriteLine("learn 50000")  
  
    newNeuralNetFilename = _  
        neuralNetFilenamePrefix & _  
        "_" & _  
        seed & _  
        ".nnd"  
    ' textFile.WriteLine("savenet " + newNeuralNetFilename)  
  
    textFile.WriteLine("set testfile Holdout.txt")  
    textFile.WriteLine("test")  
  
    newNeuralNetRecallFilename = _  
        "Holdout_" & _  
        neuralNetFilenamePrefix & _  
        "_" & _  
        seed & _  
        ".nnd"  
    ' textFile.WriteLine("savenet " + newNeuralNetRecallFilename)  
  
    textFile.WriteLine("set testfile 'Train.txt")  
    textFile.WriteLine("test")  
  
    newNeuralNetRecallFilename = _  
        "Learning_" & _  
        neuralNetFilenamePrefix & _  
        "_" & _
```

```

    seed & _
    ".nnd"

    textFile.Close

    mostRecentSeed = seed
    mostRecentNeuralNetFilenamePrefix = neuralNetFilenamePrefix

End Sub

'-----
' Run command in background.
'-----
Sub RunCommandInBackground(command)

    Set shell = WScript.CreateObject("WScript.Shell")

    command = "%comspec% /c " & command

    hideWindowAndActivateAnother = 0
    waitForProgramToFinish = True
    shell.Run command, hideWindowAndActivateAnother, waitForProgramToFinish

End Sub

'-----
' Run command in foreground.
'-----
Sub RunCommandInForeground(command)

    Set shell = WScript.CreateObject("WScript.Shell")

    activateAndDisplayWindow = 1
    dontWaitForProgramToFinish = false
    shell.Run command, activateAndDisplayWindow, dontWaitForProgramToFinish

End Sub

'-----
' Returns the directory containing NeuralWorks EXEs.
'-----
Function GetNeuralWorksInstallDirectory

    Set shell = CreateObject("WScript.Shell")

```

```

GetNeuralWorksInstallDirectory = _
    shell.ExpandEnvironmentStrings("%NEURAL_WORKS_INSTALL_DIR%")

End Function

'-----
' Runs the NeuralWorks automation script that's sitting in the current directory
' (automate.dat).
'-----
Sub RunNeuralWorksAutomationScript

    Dim commandToRunScript

    'commandToRunScript = "" & GetNeuralWorksInstallDirectory() & "\" & "nw2 -
xautomate >>" & logFilename & " 2>&1" & ""
    commandToRunScript = "" & GetNeuralWorksInstallDirectory() & "\" & "nw2 -
xautomate >>" & logFilename & " 2>&1" & ""

    RunCommandInBackground commandToRunScript
    RunCommandInBackground "type automate.dat >> " & logFilename

    Dim nnrFilename
    nnrFilename = _
        "Holdout_" & _
        mostRecentNeuralNetFilenamePrefix & _
        "_" & _
        mostRecentSeed & _
        ".nnr"

    Dim backupNnrCommand
    backupNnrCommand = "copy Holdout_txt.nnr " & nnrFilename

    RunCommandInBackground backupNnrCommand
'-----
' New section to save the training-set-pass NNR
'-----
    Dim tspFilename
    tspFilename = _
        "Learning_" & _
        mostRecentNeuralNetFilenamePrefix & _
        "_" & _
        mostRecentSeed & _
        ".nnr"

    Dim backupTspCommand

```

```

    backupTspCommand = "copy Train_txt.nnr " & tspFilename

    RunCommandInBackground backupTspCommand

End Sub

'-----
' Opens a log of the NeuralWorks automation tool's output in Window's "Notepad"
' text editor.
'-----
Sub OpenNeuralWorksLogInNotepad

    RunCommandInForeground "notepad " & logFilename

End Sub

'-----
' Display's the script's progress in completing all neural network automation.
'-----
Sub DisplayProgress(runsRemaining, totalRuns)

    WScript.Echo " " & runsRemaining & "/" & totalRuns & _
        " iterations remaining"

End Sub

'-----
' script entry point
'-----

RunCommandInBackground("del " & logFilename)

neuralNetFilenameList = GetNeuralNetFilenameList()

seedList = GetSeedList()

totalRuns = _
    (UBound(neuralNetFilenameList) + 1) * _
    (UBound(seedList) + 1)

runsRemaining = totalRuns

For Each neuralNetFilename In neuralNetFilenameList

    For Each seed In seedList

```

GenerateNeuralWorksAutomationScript neuralNetFilename, seed

RunNeuralWorksAutomationScript

DisplayProgress runsRemaining, totalRuns

runsRemaining = runsRemaining - 1

Next

Next

OpenNeuralWorksLogInNotepad

APPENDIX M: C Code for the Top-Generalizing BIC-based Network

This code was automatically generated by NeuralWare's NeuralWorks from the network created by Mehmet Vurkaç.

```
/* Fri Nov 04 07:37:30 2011 (recall.c) */
/* Header file is <recall.h> */
/* Recall-Only Run-time for <ABE-b> */
/* Control Strategy is: <backprop> */

#if defined(__STDC__) || defined(__cplusplus)
#define ARGS(x) x
#else
#define ARGS(x) ()
#endif /* __STDC__ */
#if defined(__cplusplus)
extern "C" {
#endif

/* --- External Routines --- */
extern double exp ARGS((double));

/* *** MAKE SURE TO LINK IN YOUR COMPILER's MATH LIBRARIES *** */

#if defined(STATIC_WTS)
typedef struct _pewts {
    short      sPEFlag; /* Flag for weight type */
    unsigned short usPESrc; /* index of source PE */
    float      fPEWt; /* value of weight for PE */
} PEWTS;
#define CN_VAR 0 /* variable weight */
#define CN_SET 2 /* set weight */
#define CN_MOD 3 /* mod weight */

#define ASof(x) (sizeof(x)/sizeof(x[0]))
#define SWC(x) x

static PEWTS taPEWts0032[] = {
    { CN_VAR, 1, (float)-1.209466 },
    { CN_VAR, 18, (float)0.8807662 },
    { CN_VAR, 19, (float)3.734666 },
    { CN_VAR, 20, (float)2.410166 },

```



```

        { CN_VAR, 21, (float)-2.277215 },
        { CN_VAR, 22, (float)-4.329589 },
        { CN_VAR, 23, (float)3.730004 },
        { CN_VAR, 24, (float)2.160049 },
        { CN_VAR, 25, (float)2.625435 },
        { CN_VAR, 26, (float)-5.596211 },
        { CN_VAR, 27, (float)2.23971 },
        { CN_VAR, 28, (float)-0.4534248 },
        { CN_VAR, 29, (float)-1.635675 },
        { CN_VAR, 30, (float)-2.837386 },
        { CN_VAR, 31, (float)-3.1529 }
    };
#else
#define SWC(x)
#endif /* #if defined(STATIC_WTS) */

/* --- NOTE: Network has (Yin[16]) inputs and (Yout[1]) outputs --- */
#if defined(__STDC__) || defined(__cplusplus)
int NN_Recall( const void *NetPtr, const float *Yin, float *Yout )
#else
int NN_Recall( NetPtr, Yin, Yout )
    void *NetPtr; /* Network Pointer (not used) */
    float *Yin, *Yout; /* Data In=16 Out=1 */
#endif /* __STDC__ */
{
    float Xout[33], Xsum[33]; /* work arrays */
    long ICmpT=0; /* temp for comparisons */
    SWC(PEWTS *tpPEW=0;) /* PE Weight Pointer */
    SWC(int nPEX=0;) /* PE Weight counter */
    SWC(int nFF=0;) /* PE First Flag */
    SWC(float fWt=0;) /* float work weight */

    /* *** WARNING: Code generated assuming Recall = 0 *** */
    SWC(Xout[1] = (float)1;) /* Initialize Bias */

    /* Read and scale input into network */
    Xout[2] = Yin[0];
    Xout[3] = Yin[1];
    Xout[4] = Yin[2];
    Xout[5] = Yin[3];
    Xout[6] = Yin[4];
    Xout[7] = Yin[5];
    Xout[8] = Yin[6];
    Xout[9] = Yin[7];
    Xout[10] = Yin[8];

```

```

Xout[11] = Yin[9];
Xout[12] = Yin[10];
Xout[13] = Yin[11];
Xout[14] = Yin[12];
Xout[15] = Yin[13];
Xout[16] = Yin[14];
Xout[17] = Yin[15];

```

LAB107:

```

/* Generating code for PE 0 in layer <Hidden1> (3) */
Xsum[18] = (float)(-0.81172377) + (float)(-1.6960096) * Xout[2] +
(float)(1.0204016) * Xout[3];
/* Generating code for PE 1 in layer <Hidden1> (3) */
Xsum[19] = (float)(2.7283597) + (float)(-2.661437) * Xout[2] + (float)(-6.1617775) *
Xout[4];
/* Generating code for PE 2 in layer <Hidden1> (3) */
Xsum[20] = (float)(3.3842492) + (float)(-1.5893893) * Xout[2] + (float)(1.876165) *
Xout[10] +
(float)(-3.1415153) * Xout[11];
/* Generating code for PE 3 in layer <Hidden1> (3) */
Xsum[21] = (float)(0.47418568) + (float)(3.2498701) * Xout[2] + (float)(-2.0458055)
* Xout[17];
/* Generating code for PE 4 in layer <Hidden1> (3) */
Xsum[22] = (float)(4.0053868) + (float)(-4.5400424) * Xout[3] + (float)(-4.2807059)
* Xout[5];
/* Generating code for PE 5 in layer <Hidden1> (3) */
Xsum[23] = (float)(-0.053333879) + (float)(2.3765006) * Xout[6] + (float)(-
2.4425097) * Xout[7] +
(float)(6.5449004) * Xout[8];
/* Generating code for PE 6 in layer <Hidden1> (3) */
Xsum[24] = (float)(-0.21082652) + (float)(-4.4376731) * Xout[7] +
(float)(1.4935943) * Xout[10];
/* Generating code for PE 7 in layer <Hidden1> (3) */
Xsum[25] = (float)(1.3428322) + (float)(-4.7555795) * Xout[9] + (float)(-
0.91127557) * Xout[10];
/* Generating code for PE 8 in layer <Hidden1> (3) */
Xsum[26] = (float)(-1.1912055) + (float)(-6.529182) * Xout[10] + (float)(7.4855285)
* Xout[11] +
(float)(-6.7891407) * Xout[12];
/* Generating code for PE 9 in layer <Hidden1> (3) */
Xsum[27] = (float)(0.1924693) + (float)(1.1016182) * Xout[10] + (float)(-4.436893) *
Xout[13];
/* Generating code for PE 10 in layer <Hidden1> (3) */
Xsum[28] = (float)(-0.72840422) + (float)(-0.19122335) * Xout[10] +
(float)(0.59778601) * Xout[16];

```

```

/* Generating code for PE 11 in layer <Hidden1> (3) */
Xsum[29] = (float)(-0.65263236) + (float)(2.5064108) * Xout[14];
/* Generating code for PE 12 in layer <Hidden1> (3) */
Xsum[30] = (float)(1.8148082) + (float)(-4.211267) * Xout[15] + (float)(-1.5137342)
* Xout[17];
/* Generating code for PE 13 in layer <Hidden1> (3) */
Xsum[31] = (float)(-1.4726136) + (float)(4.3973227) * Xout[16] + (float)(-
2.9602437) * Xout[17];
/* Generating code for PE 0 in layer <Hidden1> (3) */
Xout[18] = (float)(1.0 / (1.0 + exp( -Xsum[18] )));

/* Generating code for PE 1 in layer <Hidden1> (3) */
Xout[19] = (float)(1.0 / (1.0 + exp( -Xsum[19] )));

/* Generating code for PE 2 in layer <Hidden1> (3) */
Xout[20] = (float)(1.0 / (1.0 + exp( -Xsum[20] )));

/* Generating code for PE 3 in layer <Hidden1> (3) */
Xout[21] = (float)(1.0 / (1.0 + exp( -Xsum[21] )));

/* Generating code for PE 4 in layer <Hidden1> (3) */
Xout[22] = (float)(1.0 / (1.0 + exp( -Xsum[22] )));

/* Generating code for PE 5 in layer <Hidden1> (3) */
Xout[23] = (float)(1.0 / (1.0 + exp( -Xsum[23] )));

/* Generating code for PE 6 in layer <Hidden1> (3) */
Xout[24] = (float)(1.0 / (1.0 + exp( -Xsum[24] )));

/* Generating code for PE 7 in layer <Hidden1> (3) */
Xout[25] = (float)(1.0 / (1.0 + exp( -Xsum[25] )));

/* Generating code for PE 8 in layer <Hidden1> (3) */
Xout[26] = (float)(1.0 / (1.0 + exp( -Xsum[26] )));

/* Generating code for PE 9 in layer <Hidden1> (3) */
Xout[27] = (float)(1.0 / (1.0 + exp( -Xsum[27] )));

/* Generating code for PE 10 in layer <Hidden1> (3) */
Xout[28] = (float)(1.0 / (1.0 + exp( -Xsum[28] )));

/* Generating code for PE 11 in layer <Hidden1> (3) */
Xout[29] = (float)(1.0 / (1.0 + exp( -Xsum[29] )));

/* Generating code for PE 12 in layer <Hidden1> (3) */

```

```

Xout[30] = (float)(1.0 / (1.0 + exp( -Xsum[30] )));

/* Generating code for PE 13 in layer <Hidden1> (3) */
Xout[31] = (float)(1.0 / (1.0 + exp( -Xsum[31] )));

/* Generating code for PE 0 in layer <Out> (4) */
#if defined(STATIC_WTS)
for( nFF=0, nPEX=ASof(taPEWts0032), tpPEW = &taPEWts0032[0]; --nPEX >=
0; tpPEW++) {
    if ( nFF == 0 ) {
        Xsum[32] = (float)(tpPEW->fPEWt * Xout[tpPEW->usPESrc]);
        nFF = 1;
    } else {
        Xsum[32] += (float)(tpPEW->fPEWt * Xout[tpPEW->usPESrc]);
    }
}
#else /* #if defined(STATIC_WTS) */
Xsum[32] = (float)(-1.209466) + (float)(0.88076621) * Xout[18] + (float)(3.7346663)
* Xout[19] +
(float)(2.4101663) * Xout[20] + (float)(-2.2772145) * Xout[21] + (float)(-
4.3295894) * Xout[22] +
(float)(3.7300043) * Xout[23] + (float)(2.160049) * Xout[24] + (float)(2.6254351) *
Xout[25] +
(float)(-5.596211) * Xout[26] + (float)(2.2397101) * Xout[27];
Xsum[32] += (float)(-0.45342484) * Xout[28] + (float)(-1.6356754) * Xout[29] +
(float)(-2.8373857) * Xout[30] +
(float)(-3.1528997) * Xout[31];
#endif /* #if defined(STATIC_WTS) */
Xout[32] = (float)(1.0 / (1.0 + exp( -Xsum[32] )));

/* De-scale and write output from network */
Yout[0] = Xout[32];

/* Generating code for PE 0 in layer <Out> (4) */
return( 0 );
}

#if defined(__cplusplus)
}
#endif

```

APPENDIX N: Sample Data in the Firm-Teacher Context

0	0	1	1	1	1	0	0	1	1	1	1	0	0	1	0	2	1	0	0	1	0	0	1	0	0	strong
1	0	1	0	0	1	1	0	0	0	0	1	1	1	0	1	0	0	0	0	0	0	0	0	0	1	strong
0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	1	1	0	1	0	0	1	0	0	1	0	WBD
1	1	0	1	0	0	1	0	1	0	1	0	1	1	0	0	1	0	1	0	0	1	0	0	1	0	v_strong
1	0	1	0	1	1	1	0	1	0	1	1	0	0	1	1	2	1	0	0	1	0	0	1	0	0	strong
0	0	0	1	1	0	1	0	1	0	1	1	0	0	1	1	1	0	1	0	0	1	0	0	1	0	str/VS
0	1	1	1	1	1	1	1	1	1	1	0	0	0	1	0	1	0	1	0	0	1	0	0	1	0	str/VS
0	1	0	1	1	0	0	0	0	1	1	1	1	1	0	0	1	0	1	0	0	1	0	0	1	0	ave
1	0	1	0	0	0	0	0	1	0	1	1	1	0	1	1	2	1	0	0	1	0	0	1	0	0	ave
1	1	0	1	1	1	0	1	0	1	1	1	0	1	1	1	2	1	0	0	1	0	0	1	0	0	weak
0	0	0	0	0	0	0	0	0	1	1	0	1	0	0	0	2	1	0	0	1	0	0	1	0	0	weak
1	0	0	0	0	0	1	1	0	0	1	0	0	1	0	0	2	1	0	0	1	0	0	1	0	0	WBD
0	1	0	1	1	1	0	1	0	1	0	1	0	1	1	1	2	1	0	0	1	0	0	1	0	0	vwk
0	0	1	1	1	1	0	0	0	0	0	0	1	1	1	0	2	1	0	0	1	0	0	1	0	0	strong
1	1	0	1	0	0	0	0	0	0	1	0	0	0	0	0	1	0	1	0	0	1	0	0	1	0	strong
0	0	0	0	0	1	1	0	1	0	1	1	1	1	1	0	1	0	1	0	0	1	0	0	1	0	WBD
0	0	1	0	0	0	0	1	1	1	0	0	0	0	1	0	2	1	0	0	1	0	0	1	0	0	strong
0	1	1	0	1	1	1	1	0	0	1	0	0	0	0	0	2	1	0	0	1	0	0	1	0	0	weak
0	1	0	1	1	1	1	1	0	0	0	1	1	1	1	1	1	0	1	0	0	1	0	0	1	0	v_strong
1	0	1	1	0	1	1	1	1	1	0	0	1	1	1	1	2	1	0	0	1	0	0	1	0	0	str/VS
1	1	0	1	1	1	1	1	1	1	1	0	1	0	1	0	1	0	1	0	0	1	0	0	1	0	strong
0	1	1	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	ave
0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	1	0	0	1	0	vwk
1	0	1	1	0	1	1	0	1	0	0	1	0	0	1	0	2	1	0	0	1	0	0	1	0	0	ave
0	0	0	1	0	0	0	1	0	0	1	1	1	0	1	1	1	0	1	0	0	1	0	0	1	0	strong
1	0	0	0	0	1	1	1	0	0	1	0	1	0	0	1	1	0	1	0	0	1	0	0	1	0	weak
1	1	1	0	0	1	0	1	1	1	1	1	1	0	0	0	2	1	0	0	1	0	0	1	0	0	strong
1	1	1	1	0	0	1	1	1	0	1	0	1	0	0	0	1	0	1	0	0	1	0	0	1	0	weak
1	1	0	0	0	0	1	0	0	1	0	1	1	1	1	1	2	1	0	0	1	0	0	1	0	0	weak
1	0	1	0	1	0	0	0	0	0	0	0	1	1	0	0	2	1	0	0	1	0	0	1	0	0	WBD
0	0	0	0	1	1	1	1	1	0	1	0	1	0	0	1	1	0	1	0	0	1	0	0	1	0	str/VS
0	0	0	1	1	0	0	0	0	1	1	1	1	0	0	1	1	0	1	0	0	1	0	0	1	0	weak
1	1	0	1	1	0	1	0	1	0	1	0	1	0	0	1	1	0	1	0	0	1	0	0	1	0	v_strong
1	1	0	0	1	1	1	0	0	0	0	0	0	0	0	1	1	0	1	0	0	1	0	0	1	0	strong
1	0	0	0	1	0	0	0	0	0	0	0	1	1	1	0	2	1	0	0	1	0	0	1	0	0	ave

Figure 75: Sample I/O vectors followed by clave-direction class (0, 1, 2, 3) and various encodings of such. Membership degrees are in the rightmost column.

APPENDIX O: Random Seeds

3765	8391	5169	7523
1344	5795	3376	2209
7082	8522	82	1716
7659	1737	3932	9147
2448	4698	4118	3990
8788	2168	1627	6209
3395	906	7407	868
9258	7103	2758	6453
6561	4147	6319	8252
9981	920	9679	5086
973	6432	7354	9136
7053	2658	8201	7847
6074	7206	307	7619
3674	3708	751	4839
6141	1674	7062	432
4401	660	3526	5845
495	9939	1855	181
1910	5715	2711	4858
6857	4160	7755	6430
3791	8801	968	5067
6523	7496	8250	9194
6056	6223	5469	23
4703	9863	5606	8294
9115	5298	6134	5292
2563	757	8773	6150

3944	2991	4890	1025
1153	4846	5880	4327
9796	6763	2212	7742
3543	7730	1331	6298
8995	7119	5583	5323
5152	8198	1759	5195
856	2279	9318	8638
6417	98	975	1188
6631	1973	3477	9795
4326	4499	1250	8725
6843	8870	1197	3523
2560	1989	1931	786
2467	1670	9874	5037
3400	3762	5705	3661
9016	8084	2900	2035
8469	2597	2703	359
5269	4024	1281	2230
5488	8271	3285	563
2023	3086	8189	8187
4321	5452	2539	5053
5955	1438	545	2293
9458	1798	4639	3605
2550	793	2320	561
2065	8847	239	3350
8276	2882	9629	3082
7111	2675	4733	4020

474	4658	6556	1479
6132	6485	1429	8297
2855	3704	4138	8897
8737	3353	6622	837
448	7233	1227	1464
9859	8260	2708	2605
3883	7279	5791	3109
2607	2827	3628	167
2906	9848	4501	7511
75	2163	3127	4625
323	1622	7113	1231
7117	5224	3466	3697
7016	3531	4217	2717
9484	6104	7364	1740
7631	3801	1371	3130
9409	880	8092	6450
3761	3074	8339	9903
8854	3767	747	2796
779	3134	1120	3476
3081	1185	7011	3819
9539	7941	5243	68
7650	3843	2191	5039
3779	4820	5801	7100
7350	256	6941	7716
5534	7656	180	4744
6071	9244	7873	6526

9875	501	4943	9137
3699	6913	9186	1383
8529	5811	592	297
7538	4025	126	3071
9327	6389	3931	4495
2503	6068	1450	5153
5273	2774	8955	4819
5769	9953	1294	3869
8196	3259	5017	4918
5805	9138	7150	2559
3936	6126	2851	5031
1968	1164	7915	8571
67	4421	3232	4396
7648	8563	9285	3715
4170	273	4173	7445
2890	7173	6268	7157
6426	9845	7764	7289
2425	8205	939	2337
9882	934	6894	1521
7222	9634	7860	
6131	4339	9253	
3318	9765	9594	