5-2019

Attrition Effects in Mandarin-English Bilinguals of Varying Proficiencies

Sarah Charlotte Elkinton
Portland State University

Recommended Citation

Elkinton, Sarah Charlotte, "Attrition Effects in Mandarin-English Bilinguals of Varying Proficiencies" (2019). *University Honors Theses.* Paper 671.
https://doi.org/10.15760/honors.685

Attrition Effects in Mandarin English Bilinguals of Varying Proficiencies

By
Sarah Charlotte Elkinton

An undergraduate honors thesis submitted in partial fulfillment of the requirements for
the degree of
Bachelor's of Science
In
University Honors
And
Speech and Hearing Sciences

Thesis Advisor

Carolyn Quam, PhD
2019

Attrition Effects in Mandarin-English Bilinguals of Varying Proficiencies

Sarah Elkinton
Portland State University

Abstract

This thesis is a continuation and expansion of Quam and Creel (2017a), who found that English-dominant native Mandarin speakers showed an attrition effect in their ability to process tones. Experiment 1 of this thesis aimed to replicate their findings, which were mostly upheld. Experiment 2 aimed to differentiate two alternative explanations for the selective tone attrition effects offered in Quam and Creel (2017a). One explanation is that lexical tones are more prone to attrition because of their inherent properties. A second explanation is that consistent use of English vowel categories reinforced the particular Mandarin vowel contrasts that were used, as opposed to the tones, which do not occur in English. We aimed to differentiate these explanations by testing Mandarin-English bilinguals on vowel contrasts that do not occur in English. Continued recruitment efforts will enable us to differentiate the two explanations.

Acnowlagements and Thank You's

Table of Contents

List of Figures and Tables

Introduction

In the United States of America, the number of bilingual speakers is on the rise, bringing the United States in line with other countries around the world (Kohnert, 2013; US Census Bureau, n.d.). Worldwide it is estimated that half the population speaks two languages, and approximately 18.5% of speakers in the United States have a first language other than English (Kohnert, 2013; US Census Bureau, n.d.). For these speakers, as soon as they interact with people who speak a different language, they become functional bilinguals - individuals who need two or more languages in their daily lives (Kohnert, 2013). Despite the growing number of bilingual speakers in the U.S. and worldwide, there is a lack of understanding of how bilingual speakers store the phonemic inventories of their respective languages, and more importantly, how the sound systems of the two languages interact. This is particularly true when one language uses a feature or acoustic dimension that the other does not.

This study continues a line of research begun by Quam and Creel (2017a), who found that native Mandarin speakers who have become English dominant show attrition in their ability to access lexical tone information. Quam and Creel offered two explanations for their finding. One is that this attrition effect was due to how the brain stores and accesses lexical tone information. The second is that a general lack of use of Mandarin caused overall attrition to both tone and vowels, but English-dominant listeners could recruit English vowel categories to process the particular Mandarin vowels that were used. English dominant bilinguals were unable to recruit English lexical tone categories in a similar manner because English is a non-tonal language.

Lexical tones

Lexical tone is contrastive in some languages, like Mandarin, but not in others, like English. In tonal languages, lexical tone is a pitch pattern that changes the meaning of the word. For example, Mandarin uses four lexical tones (tone 1-high flat; tone 2-rising; tone 3-dipping; tone 4-falling). This means that /mɑ/ (*ma*) changes its meaning depending on which tonal pattern is used. The tonal pattern is indicated by a number after the word, so *ma*1 is said on a high flat tone (and means 'mother') while *ma*4 is said on a high to low falling tone (and means 'to scold'). English does not use tone at the lexical level, so /k^p/ (cup) always refer to the drinking vessel. It can be said with a raising inflection ("cup?") or a falling inflection ("cup."), but either inflection refers to the same object: a drinking vessel.

Quam and Creel (2017a) are not the first to suggest that lexical tone is stored separately and accessed differently in the brain than segments (the consonants and vowels of a language). Burnham et al. (2011) found that in Thai, Cantonese and Australian English speakers, tonal awareness (metalinguistic awareness of lexical tones) lagged behind segmental awareness (metalinguistic awareness of the consonants and vowels of a word) at every education level tested, up to adults with post-secondary education. This lag in the development of tone processing and perception is not limited to those developing their knowledge of a tonal language (Burnham et al., 2011). Previous studies have found that Mandarin speakers take longer to make judgments of tonal differences than segmental differences (usually vowel differences) in a variety of tasks and tests. The delayed response for tone-disambiguated trials when compared to segment-differentiated trials indicate that

Mandarin speakers take more time to access the tonal component of a word than to access the segmental components of a word (Taft & Chen, 1992; Ye & Connine, 1999). Burnham et al's (2011) study of the development of segmental awareness compared to tonal awareness in Thai speakers, when taken in conjunction with research studies on tonal processing that indicate lexical tones are accessed along a separate, slower time course than segmental information indicate there is something unique about how the mind represents and accesses lexical tone as compared to consonants and vowels (Cheung, Chen, Lai, Wong, & Hills, 2002; Taft & Chen, 1992; Ye & Connine, 1999). However, other evidence suggests more comparable processing of tones and segments. Malins and Joanisse (2010) asked whether Mandarin speakers access tones and segments in a similar time course during spoken-language processing. In a familiar-word identification task, they asked Mandarin speakers to select an item from an array of items that were either tonally or segmentally related and unrelated. When Mandarin speakers heard a target word and tried to identify the picture it referred to, in tone-disambiguated trials they looked at tone-disambiguated competitors no more frequently than they looked to segmentally-disambiguated or unrelated distractors. The same was true in segmentally-disambiguated trials. However, a methodological issue with the study could have minimized differences between trial types. Target pictures appeared many more times across trials than competitors. This over-representation of target pictures likely minimized looks to competitors, reducing differences between trial types. In an event-related potential (ERP) study of neural responses to speech, Schirmer, Tang, Penney, Gunter, and Chen, (2005) found that altering the rime (the central vowel and possibly the last consonant of a word) of an anticipated word created the same

N400 ERP (the "confusion" response) as altering the tone of the anticipated word,

implying that tones play just as important a role in identifying words in a speech stream

as segments. Taken together, Malins and Joanisse (2010) and Schirmer et al. (2005)

challenge the first explanation put forward in Quam and Creel (2017a): that lexical tone

is more prone to attrition because of how the brain stores and accesses this information.

Perception and Perceptual Categories

The second explanation put forward by Quam and Creel to explain their finding of

attrition is that English-dominant Mandarin speakers were able to recruit English vowel

perceptual categories to process vowels in Mandarin (Quam and Creel, 2017a). Speech

perception, which involves identifying the sounds one is hearing, is central to this

explanation. Werker (2018) argues that perception of sound is key to the development

of language. At birth, infants are able to perceive all the sounds of the world's

languages, but lose that ability as they age. This perceptual narrowing enables the

infant to become a specialist in the sounds of their first language (L1). However, this

perceptual narrowing trajectory is more complex for bilinguals. One reason is that

bilinguals must mentally represent two different sets of sound categories. A second

reason is that, for bilinguals, these perceptual categories are not static over time, even

beyond the first year of life, when monolingual infants have already zeroed in on native-

language sound categories (Bosch & Sebastián-Gallés, 2003; Polka & Werker, 1994;

Werker, 2018) The amount and quality of language inputs in each language are

continually changing their balance in bilingual speakers. There is good evidence to

suggest an individual's languages are in a constant state of flux. As far back as 1989,

scholars were arguing for a dynamic view of bilingual's languages (Grosjean, 1989).

Grosjean argued that bilingual individuals were not simply two monolinguals in the same body, but are complete, capable, complex and dynamic listeners and speakers of both their languages and are able to tailor their language use to their specific situations. This tailoring occurs at the lexical level and at the phonemic level, with bilingual speakers suppressing the non-desired language, or even changing phonemic inventories based on input. Such changes have been documented in children and adults after surprisingly minimal exposure periods (Chang, 2013; Gildersleeve-Neumann, Peña, Davis, & Kester, 2009).

The Perceptual Assimilation Model (PAM) posits that sound perception in a second language (L2) is driven by the sound maps of a speaker's first language (L1) (Best, Goldstein, Tyler, & Nam, 2009; Best, McRoberts, & Goodell, 2001). This theory posits that as individuals learn their first language, they are attending to two separate dimensions, the *phonological distinctiveness* (the unique articulations needed to produce a sound) and the *phonological consistency* (the changes of articulation that do not change meaning) of the phonemes of their first language (L1). In the context of learning a second language, PAM posits that sounds of a new language are categorized by how similar or different they are to the sounds of L1 (Tyler et al., 2014). L2 sounds that are perceived as speech are "categorized" if they have an identifiable relationship with a sound in L1, and are considered "uncategorized" if they do not. Considering two sounds that form a contrast in L2, a particular sound pair can be (1) "two category" sounds, where each L2 sound maps onto separate L1 perceptual categories; (2) "category goodness" sounds, where two sounds in L2 share one perceptual category in L1, but one is a 'good' fit while the other is noticeably less prototypical for the L1

category; or (3) "single category" sounds, where two L2 sounds share a single L1

categorization equally (See Reid, et al. (2015) or Tyler, Nil, Best, Faber & Levitt (2014)

for a more complete description). PAM predicts that L2 sounds that are "two category"

phonemes will be easiest for an L1 speaker to discriminate and identify, while "category

goodness" sounds will be slightly less easy to identify; and those L2 sounds that are

"single category" will be the hardest for L1 individuals to identify.

In the present study, since we are most interested in how individuals' L2

perceptual categories impact processing of sounds in their L1, we are applying PAM not

from L1 to L2, but from L2 to L1. Experiment 2 of this thesis, in particular, explores this

"backwards" application of PAM, represented by the second of Quam and Creel's

hypotheses, that English-dominant Mandarin bilinguals recruit their English vowel

categories to process Mandarin vowels. If this is the case, we expect to see that two

Mandarin phonemes that map to a single English phoneme category will be the most

difficult for English-dominant bilingual speakers to classify. By contrast, we expect to

see that two Mandarin phonemes that map onto two distinct English phonemes will be

easiest for English-dominant bilinguals to process.

The present experiments

There are two experiments in this thesis. Experiment 1 is a replication of Quam

and Creel (2017a). This experiment uses the same methods as the 2017 study but tests

a new group of participants (a group that participated in Experiment 3 of Quam & Creel,

2017b prior to completing Experiment 1). The goal of Experiment 1 is to test the

robustness of the findings of Quam and Creel (2017a).

Experiment 2 is designed to differentiate two potential explanations for the finding of selective attrition for tone in Quam and Creel (2017a). The first explanation (for which we have just reviewed evidence from the literature, above) is that lexical tones are more prone to attrition because of their inherent properties. A second explanation, based on PAM, is that consistent use of English vowel categories reinforced the particular Mandarin vowel contrasts that were used (which were likely "two category" contrasts), as opposed to the tones, which do not occur in English. Regarding the second explanation, Quam and Creel (2017a) compared processing of tones versus vowels, but did not systematically consider the relationships between L1 (Mandarin) and L2 (English) vowel categories for the Mandarin vowel contrasts in the stimuli. Experiment 2 builds on the work of Quam and Creel by intentionally including a wider variety of Mandarin vowels (in terms of their assimilability to English categories). This enables us to confirm or disconfirm explanation 2, that the English-dominant participants tested by Quam and Creel were able to recruit the perceptual categories of English to process the Mandarin vowels. If we were to disconfirm explanation 2, this would provide indirect evidence for explanation 1.

Experiment 1

Methods

Participants.

Participants for this experiment were 61 Mandarin-English bilingual adults (37 women, mean age=20 years old, SD= 1 year, range 18-23 years old). Three participants were excluded because their language background included as much or more exposure to other tonal languages or dialects as Mandarin. Five participants were

included in reaction time and accuracy analysis but excluded from gaze data analysis:

four due to missing files, and one for failure to achieve a target plus competitor fixation

of 80% in the gaze data (Quam & Creel, 2017a,b)[1]. Prior to completing this experiment,

participants had completed an experiment investigating tone processing (Quam & Creel,

2017b, Experiment 3). Primary recruitment methods were the research participant pool

run by the Psychology and Cognitive Science Departments at the University of

California San Diego, as well as flyers posted around campus. This experiment was

approved by the Internal Review Board of the University of California San Diego. Prior

to the study, written consent was received from all participants.

Language Dominance.

Participants' language dominance was assessed in three ways: the Multilingual

Naming Test (MINT; Gollan, Weissberger, Runnqvist, Montoya, & Cera, 2012), a survey

based on the Bilingual Dominance Scale (Dunn & Fox Tree, 2009), and the Age of

Arrival in an English-dominant society. In the Multilingual Naming Test, participants are

asked to name a series of pictures, in both English and Mandarin (language order

counterbalanced). The pictures begin with familiar words, like "sun" and "hand" and

move towards less common words, ending with words like "porthole" and "axle." The

summary in Table 1 shows that English scores were both higher than Mandarin MINT

scores (paired $t(57)= 5.43$, i$<.001$) and less variable than Mandarin MINT Scores (Var

Mandarin $=146.25$, Var English$=45.97$; $F(57)=.31$, $p<.001$). English and Mandarin

scores were negatively correlated: see Figure 1.

---

[1] Gaze fixation was calculated across all trial types, familiar and novel words (Experiment 3 in Quam and Creel 2017b).

MINT scores
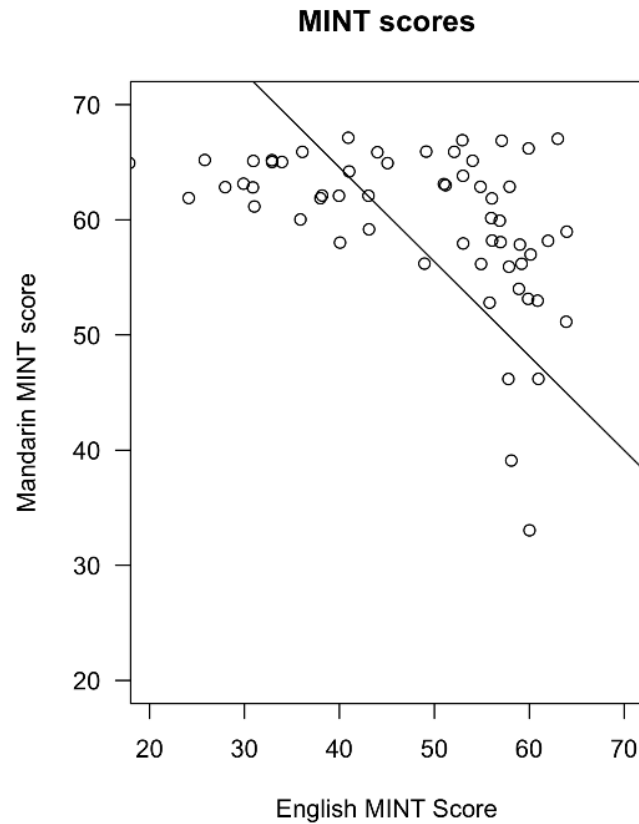
Figure 1: Multilingual Naming Test scores in Mandarin and English. Scores in both languages are negatively correlated with each other.

The Bilingual Dominance Scale is a survey of language history and present use, asking questions such as "At what age did you first learn each language?", "What language do you do math in?" and "What language do you speak with your family?" Following the instructions in Dunn and Fox Tree (2009), these questions are used to award points to each language. For both the Bilingual Dominance Scale and the Multilingual Naming Task the English score was subtracted from the Mandarin score. For these composite scores, a positive number indicates Mandarin dominance and a negative number English dominance, with larger absolute values corresponding to stronger language dominance.

The third measure was the self-reported Age of Arrival in an English-dominant society. Table 1 is a summary of the three measures of language dominance for Experiment 1 participants. As shown in Table 2, all of these measures are strongly correlated with each other. These measures are also strongly correlated with a principal component derived from all three measures via Principal Components Analysis (PCA).

Table 1: Measures of Language Dominance. Negative numbers indicate English dominance, positive numbers indicate Mandarin dominance and the absolute value indicates the magnitude of the language imbalance. Participants were English dominant on average.

	Mean *(SD)*	Range
Multilingual Naming Test		
Mandarin	48.24 *(12.09)*	18-67
English	59.91 *(6.78)*	33-67
Composite	-11.67 *(16.36)*	-47-27
Bilingual Dominance Scale	-5.34 *(11.59)*	-28-16
Age of Arrival (years)	7.59 *(7.11)*	0-19

Experimental design
   Stimuli.

Experimental design and stimuli for this experiment were taken directly from Quam and Creel (2017a). Stimuli were grouped into sets of four common words with the same initial sound: see table 3 in Quam and Creel (2017a) and Appendix B for a complete list of stimuli. Figure 2 shows a sample quadruplet comprised of *chuang2, chuang1, cha2* and *cha1. Chuang2* and *chuang1* differ only in tone and are referred to

as tone-differentiated trials, while *chuang2 and cha2* diverge at the vowel and are referred to as vowel-differentiated trials. The stimuli pairs *chuang1-cha2* and *cha1-chuang2* differ in both tone and vowel and are referred to as tone&vowel differentiated trials.

cha1   cha2

chuang1   chuang2

Figure 2: A Quadruplet from Quam and Creel 2017a. The stimuli were presented in pairs that differed in tone, vowel, or tone & vowel.

In addition to selecting quadruplets that met the above patterns, inclusion was also based on how well a word could be conveyed as an image (its imageability) and a measure of how often the characters appeared in writing (based on the Modern Chinese

Character Frequency List, Da 2004, 2005). Efforts were made to select common words as represented by their frequency in the corpus (all stimuli are in the top 33.6% of characters listed). The frequency ratios of less frequent word to more frequent word was comparable across trial types (.38:1 for tone-disambiguated and tone & vowel-disambiguated trials; .37:1 for vowel trials). Stimuli were presented equally often as target and competitor, and each word was presented once in each trial type.

These stimulus words were embedded into the carrier phrase "*Qing3 xuan3* [target]" or "Please choose [target]" recorded naturalistically by a native Mandarin speaker. Since the speaker was slightly English dominant, other Mandarin-English bilinguals (n=14) were asked to rate her pronunciation. Listeners rated this speaker as more likely to come from China or Taiwan than the United States (paired t(13) =5.37, p<.001). The same 14 listeners also rated her accent as equivalent to a very Mandarin-dominant speaker from mainland China (t(13) =.07 p=.94).

Stimuli words were matched with clip art from Microsoft Office Online repository or images from creative commons licensed photos from the website Flickr.

Table 2: Correlations for the various measures of Language Dominance. Measures of language dominance were correlated with each other, and with a Principal Component Analysis (PCA) score computed from the Multilingual Naming Test (MINT) Composite score, the Bilingual Dominance Scale (BDS) and the Age of Arrival (AOA).

	Mint Mandarin	MINT English	MINT Composite	BDS	AoA	PCA
MINT Mandarin		-.46	.93	.76	.68	.84
MINT English			-.75	-.67	-.71	-.75
MINT Composite				.84	.79	.96

| BDS | | | | | .86 | .96 |
| AOA | | | | | | .94 |

All correlations had p<.001, when α=.05

Data Analysis

Accuracy.

Participants clicked on the target picture with very high accuracy in all trial types (Tone & vowel-disambiguated M=99.87% SD=.66%; vowel-disambiguated M=99.38%, SD=2.23%; tone-disambiguated M=98.21%, SD=3.23%). In the trials differentiated by vowel & tone the lowest score was 96.43%. The lowest accuracy score for any participant in vowel disambiguated trials was 85.81%. The lowest score in tone disambiguated trials was 89.21%. These are similar accuracy results to the original Quam and Creel (2017a) paper.

To perform statistical analysis, empirical logistic (e-logit) transformation was used on the raw accuracy scores to give these potentially bounded scores an approximately normal distribution, following the best practice guidelines established in Barr (2008) and the methodology used in Quam and Creel (2017a). Analyses of variance (ANOVA) by subject and item of performance across trial type (tone&vowel, vowel, and tone disambiguated) showed that trial type was a significant predictor of accuracy (F_1(2,55) =10.653, p<.001; F_2(2,54) =14.604, p<.001). Paired t-tests showed no significant difference between trials that were tone&vowel disambiguated (M=99.88%, SD=.66%) and vowel-disambiguated trials (M=99.38%, SD =2.23%; t_1(57) =1.592, p=.117; t_2 (27)=2.364, p=.026). However, tone-disambiguated trials (M=98.21%, SD=3.23%) had

significantly lower accuracy than tone&vowel-disambiguated trials (M=99.87%,

SD=.66%; t_1(57)=4.176, p<.001; t_2(27)=3.022, p=.005). Tone-disambiguated trials were

also significantly less accurate than vowel-disambiguated trials (M=99.38%, SD=2.23%;

t_1(57)=2.886, p=.006; t_2(27)=5.138, p<.001). Quam and Creel (2017a) found similar

results, although their planned comparisons were significant in all conditions for by-

subject and by item analysis (i.e., tone&vowel trials differed from vowel trials).

In order to address the role of language dominance, the MINT Dominance, BDS

and Age of Arrival scores were combined using Principal Components Analysis (PCA).

This yielded an acceptable Kaiser-Meyer-Olkin measure of Sampling adequacy of

.757[2]. See Table 2 for individual comparisons of the PCA to individual measures of

language dominance. The PCA component was highly correlated with all the individual

measures of language dominance. The PCA for language dominance was used in an

analysis of covariance (ANCOVA) testing trial type (tone, vowel, tone&vowel) as the

within-subject categorical predictor and Mandarin dominance as a between subject

continuous predictor. This yielded a significant main effect of trial type (F(2,55=11.890,

p<.001) and language dominance (F(1,55) =7.884, p=.007), as well as a significant

interaction between trial type and language dominance (F(2,55) =7.619, p=.001).

Further investigation of the main effect of trial type revealed that tone&vowel-

disambiguated trials were not significantly different from vowel-disambiguated trials

(t(57)=1.592, p=.117), but that tone&vowel-disambiguated trials were significantly more

accurate than tone-disambiguated trials (t(57)=4.176, p<.001). Vowel-disambiguated

---

[2] A Kaiser-Meyer-Olkin score is a way of determining how representative a number is. KMO analysis returns a value between 0 and 1. Kaiser called anything below .5 "unacceptable" and anything above a .8 "meritorious"(Kaiser, 1974)

trials were also significantly more accurate than tone-disambiguated trials ($t(57)=2.886$,

$p=.006$). The main effect of language dominance indicated that more Mandarin

dominant participants achieving higher accuracy than their English-dominant peers. In a

Pearson's correlation test testing the interaction between trial type and language

dominance, clicking accuracy did not reach significance in tone&vowel differentiated

trials or vowel-differentiated trials ($r<.04$, $p>.5$). Language dominance was significantly

related to clicking accuracy in tone disambiguated trials ($r=.426$, $p=.001$). This

corresponds to what Quam and Creel (2017a) found in their study. See Figure 3 for

accuracy graphs in all trial types.
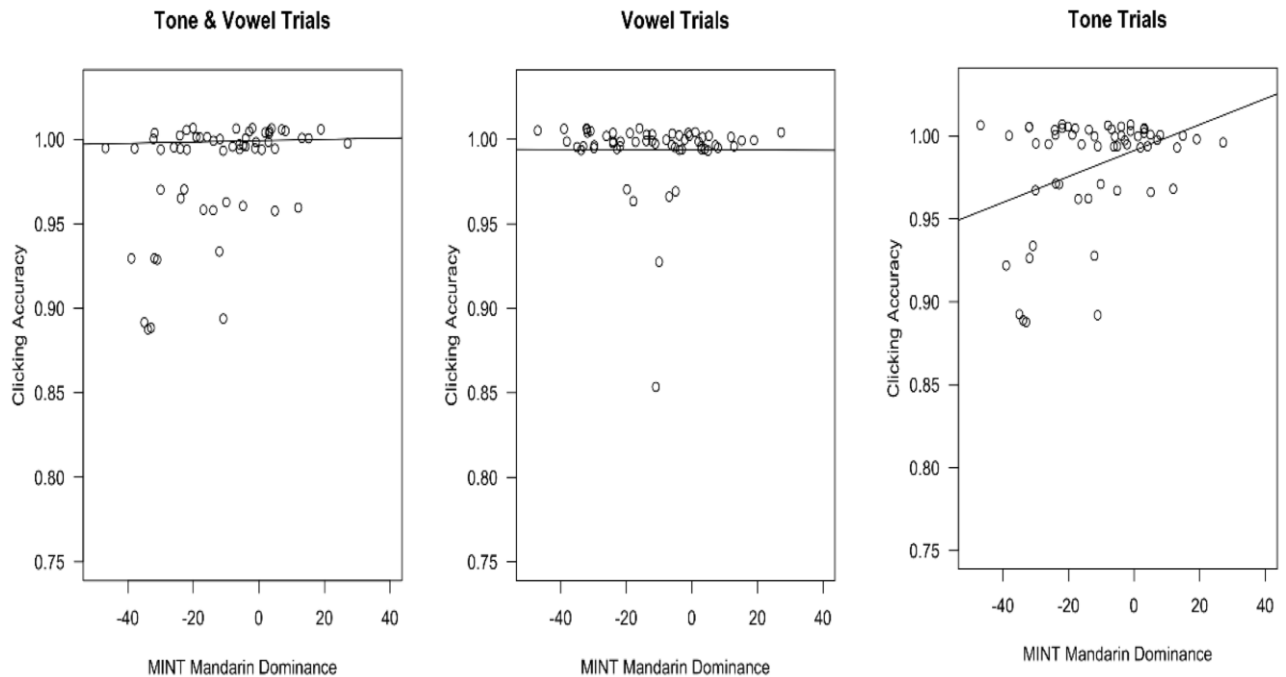
Figure 3: Accuracy vs. MINT dominance score. Accuracy was significantly correlated with Mandarin Dominance only in tone-disambiguated trials.

Reaction Time.

Reaction time (RT) analysis provides insight not only into participant accuracy but participant processing time. Additionally, because overall accuracy rates were so high, this behavioral data might distinguish more Mandarin-dominant participants from their English-dominant peers, as a more Mandarin-dominant speaker would require less time to achieve a high level of accuracy than a less Mandarin-dominant speaker would. Following the procedure of Quam and Creel (2017a), we first excluded trials that were more than three standard deviations away from a participant's mean. This resulted in an exclusion of 1.47% of trials. Reaction-time data were first analyzed with trial type (tone&vowel, vowel, tone) as the within-subject categorical predictor in ANOVAs by subject and by item. Trial type was a significant predictor of reaction times by subject and by item ($F_1(2,56)=13.495$, $p<.001$; $F_2(2,26)=118.215$, $p<.001$). Follow up t tests by subject revealed that tone&vowel-differentiated trials ($M_1=.98$ sec, $SD_1=.20$ sec; $M_2=1.74$ sec, $SD_2=.11$ sec;) were not significantly different from vowel-differentiated trials by subject ($M_1=.97$ sec, $SD_1=.16$ sec; $M_2=1.72$ sec, $SD_2=.2$ sec; $t_1(57)=.189$, $p=.851$; $t_2(27)=1.292$, $p=.207$). Tone-differentiated trials by subject and item ($M_1=1.06$ sec, $SD_1=.21$ sec; $M_2=1.81$ sec, $SD_2=.11$ sec) were significantly slower than both tone&vowel-differentiated trials and vowel-differentiated trials ($t_1(57)=-4.76$, $p<.001$; $t_2(57)=-4.08$, $p<.001$).

To assess the role of language dominance on RT, an ANCOVA was performed with trial type (tone&vowel, vowel, tone) as a within subject predictor and the language dominance PCA as the between-subject variable. This revealed a significant main effect of trial type ($F(2,55)=14.426$, $p<.001$) and not quite significant main effect of language

dominance (F(1,55)=4.58, p=.037). There was also a significant interaction of trial type

and language dominance (F(2,55)=4.932, p=.009). Follow up correlation testing

revealed that language dominance was only correlated with a faster reaction time in

tone-differentiated trials (r=.35, p=.007). Correlation between language dominance and

reaction time approached significance in tone&vowel-disambiguated trials (r=-.245,

p=.063). There was no significant correlation between reaction times in vowel-

disambiguated trials and language dominance (r=-.111, p=.407). The finding of a

correlation only in tone-disambiguated trials mirrored the findings of Quam and Creel

(2017a). Figure 4 shows reaction time plotted against language dominance.
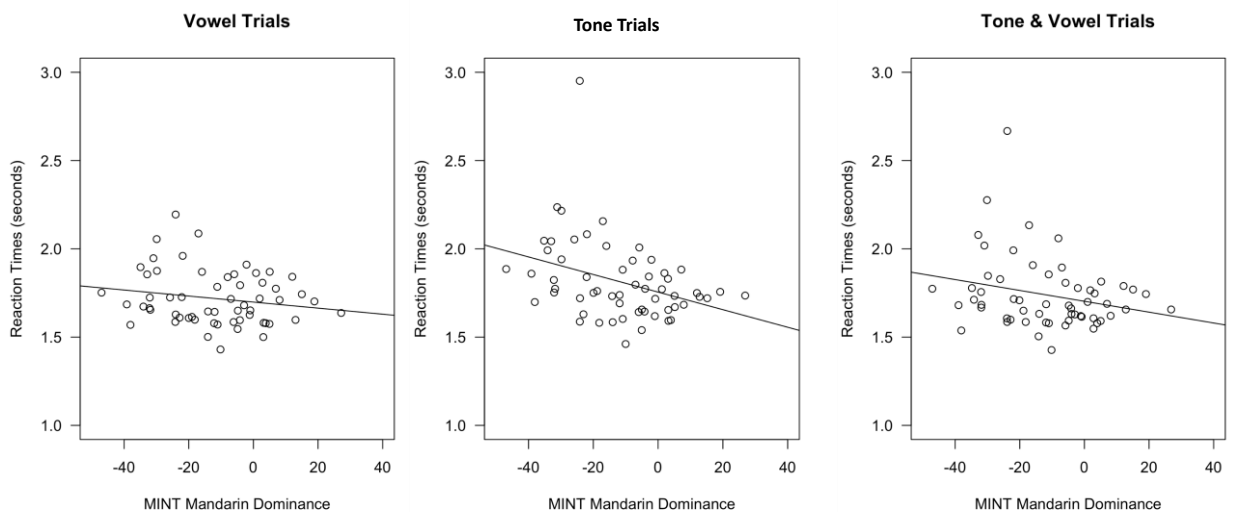
Figure 4: Reaction time vs MINT dominance Score. Reaction time was significantly correlated
with Mandarin Dominance only in Tone-disambiguated trials

The target stimuli were recorded naturalistically in the carrier phrase "*Qing3*

xuan3 [target]" or "Please choose [target]". Because of tone sandhi rules in Mandarin,

when a tone 3 word is followed by another tone 3 word, the first tone 3 is pronounced

more like a tone 2 (Wang & Li 1976). Thus, in tone-disambiguated trials, any time either

the target or competitor had a tone 3, a participant could identify the target more rapidly

based on whether the Tone 3 Sandhi rule was being applied to the 'xuan3' in the carrier phrase, something that Quam and Creel (2017) investigated in their original paper. To test whether participants were able to utilize the tone sandhi information in the carrier phrase, an ANCOVA was conducted on tone-disambiguated trials with tone type comparison (tone3-other tone) as the within subject predictor, and the language dominance PCA as a covariate. This revealed a main effect for language dominance ($F(1,56) = 8.061$, $p = .006$) but not tone type ($F(1,56) = 2.532$, $p = .117$). The interaction of language dominance and tone type was also not significant ($F(1,56) = 2.831$, $p = .098$). This non-significance of interaction effect mirrors what was reported in Quam and Creel (2017a).

Gaze Analysis.

Gaze data were limited to 50 participants (see Participants for descriptions of gaze exclusions). Like reaction time, gaze analysis allows further investigation into the time course of participants' responses, revealing more incremental processing than accuracy data can indicate.

The analysis window was limited to 200-800 milliseconds after the stimuli were presented. This was done in Quam and Creel (2017a) to reflect the time from when planned movement is first possible to the accuracy asymptote while minimizing the amount of trials extended (a total of 7.18% of trials was extended in the analysis window). Again, following the best practices recommendations in Barr (2008) and procedures in Quam and Creel (2017a), the gaze data underwent e-logit transformation to enable statistical analysis. The e-logit transformation was done to target fixation (how often the participant looked at the target picture) and competitor fixations (how often the

participant looked at the competitor picture) separately. The target advantage was computed by subtracting e-logit-transformed looks to the competitor from e-logit-transformed looks to the target. For the sake of ease of interpretation in this paper, means for target advantage and depictions in figures are untransformed, where 0 is chance looking and 1 is looking exclusively at the target.

ANOVAs were performed on target advantage with comparing trial type as a predictor, by-subject and by-item. Trial type was a significant predictor of target advantage ($F_1(2,98) = 31.234$ $p < .001$; $F_2(2,54) = 5.833$, $p = .005$). Planned analyses by subject and item indicated that target advantage was significantly greater in tone&vowel-disambiguated trials ($M = .56$, $SD = .13$) than tone-disambiguated trials ($M = .45$, $SD = .11$; $t_1(49) = 6.951$, $p < .001$; $t_2(27) = 2.891$, $p = .007$). Vowel-disambiguated trials also enjoyed greater target advantage ($M_1 = .56$, $SD = .10$) than tone-disambiguated trials ($t_1(49) = 6.668$, $p < .001$; $t_2(27) = 2.418$, $p = .023$). Tone&vowel and vowel trials were not significantly different ($t(49) = .694$, $p = .491$). This replicated what Quam and Creel (2017a) found in their study.

To test for effects of language dominance, an ANCOVA was performed with trial type (tone&vowel, vowel, tone) as the within-subject factor and language dominance as indicated by the PCA as the between-subject factor. This revealed a significant main effect for trial type ($F(2,47) = 32.966$, $p < .001$), with target advantage being significantly lower in tone-disambiguated trials than tone&vowel- or vowel-disambiguated trials. There was also a significant main effect of language dominance ($F(1,57) = 17.212$, $p < .001$), indicating that Mandarin-dominant individuals had higher target advantage across trial type than their English-dominant peers, as well as a significant interaction

effect of trial type and language dominance ($F(2,57) = 8.195$, $p = .006$). Correlation

analyses revealed that language dominance was significantly correlated with target

advantage in tone&vowel trials ($r = .362$, $p=.01$) and tone trials ($r = .559$, $p < .001$) but

not vowel trials ($r = .269$, $p = .059$). Figure 5 graphs target advantage and Mandarin

dominance in each trial type. This follows the pattern Quam and Creel (2017a)

described.
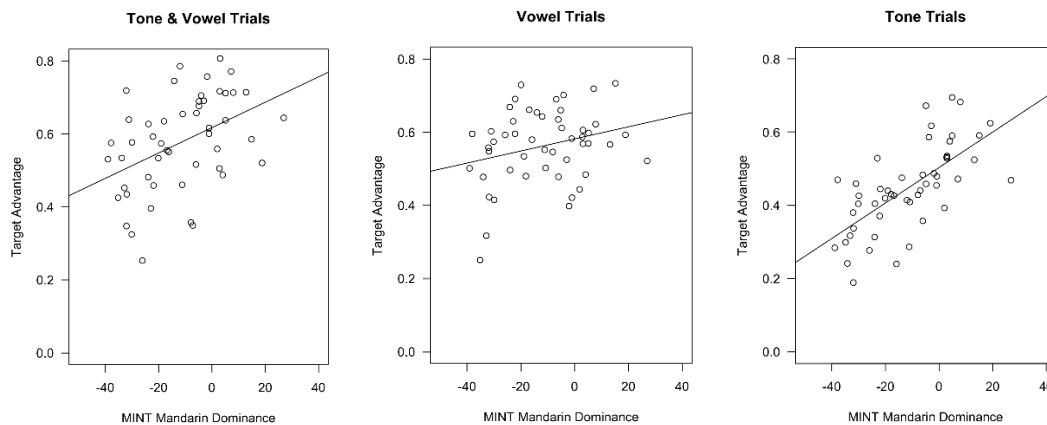
Figure 5: Target Advantage vs. Mandarin Dominance by trial type. 1 means looking exclusively at target. Mandarin dominance was significantly correlated with looks to target in tone&vowel-disambiguated trials and tone-disambiguated trials.

Since, like RT, gaze data is also time based, a tone 3 Sandhi analysis was

performed, comparing trials where either the target or competitor had a tone 3 to those

that did not have tone 3. An ANCOVA with tone type (tone3-not tone3) as the within-

subject factor and the language dominance PCA as the between-subjects factor

revealed a significant main effect for tone type ($F(1,55) = 52.400$, $p < .001$). There was

also a significant main effect of language dominance ($F(1,55) = 500.553$, $p < .001$).

There was also a significant interaction effect of tone type and language dominance

($F(1,55) = 5.101$, $p = .028$). Correlation testing revealed that both tone 3 ($r = .523$, $p <$

.001) and non-tone 3 trials (r=.367, p=.008) were positively correlated with proficiency. The interaction reflected that the association was significantly stronger in tone 3 trials. This stands in contrast to Quam and Creel (2017a) who found that the association with proficiency was only significant in Tone 3 trials.

Both tone 3 and non-tone 3 trials were correlated with proficiency, so a second investigation was done. A median-split was conducted with the PCA scores, assigning participants to "High" or "Low" proficiency Mandarin scores. This new Mandarin Dominance Category variable was used in an ANOVA as a between-subject categorical predictor, with tone type (tone 3 or non-tone 3) as the within-subject categorical variable. This revealed a significant effect of tone type (F(1,49) = 54.214, $p < .001$) and Mandarin Dominance Category (High vs Low; F(1,49) = 15.926, $p < .001$). These results mean that tone type and Mandarin dominance were significant predictors of target advantage. Planned comparisons showed that high and low proficiency Mandarin speakers were not significantly different in non-tone 3 differentiated trials (Independent equal means, equal variance assumed t(49) = 1.514, $p = .137$). However, in tone 3 trials, an independent t-test of equal means indicated that those in the "high" Mandarin-dominant category had a much higher target advantage ($M = .6680$, $SD = .14667$) than those in the "low" Mandarin dominant category ($M = .4940$, $SD = .17407$; t(49) = 3.873, $p < .001$). This shows that the more Mandarin-dominant participants were better able to take advantage of the sandhi information.

Discussion

This study was designed and conducted using the experimental procedures and methods found in Quam and Creel (2017a). It is intended to serve as a test of

replication for the finding of tone attrition in the more English-dominant bilingual

Mandarin speakers that was found in Quam and Creel (2017a). This study supports that

finding. In all three measures (accuracy, reaction time and gaze fixation), the pattern of

participant responses was fairly consistent (other than a handful of minor differences)

between this experiment and Quam and Creel 2017a.

Quam and Creel (2017a) proposed that increased exposure to a non-tonal

language (English) changes how native speakers of a tonal language (Mandarin)

process tone. Since tone is non-lexical in English, a Mandarin speaker must inhibit the

analysis of tone at the word level in order to more effectively process English. Tonal

information is generally considered to be more difficult to access than other phonetic

information in a word (See the Introduction for a summary of prior literature), and thus

potentially more prone to attrition effects, like the ones reported by Quam and Creel

(2017a).

The other explanation Quam and Creel (2017a) put forward for the selective tone

attrition effects replicated here was that the English-dominant individuals were able to

recruit the vowel categories of their dominant language (English) to process the

Mandarin vowels, masking any attrition effects that might have been seen in vowel

differentiated trials. Experiment 2 attempts to differentiate those hypotheses by more

carefully selecting vowels to include some vowels that are more comparable to tones in

their assimilability (or lack thereof) to English categories.

In addition to tone, Mandarin uses some vowel contrasts that do not exist in

English. For example, Mandarin has a vowel that is usually written in pinyin as "ü",

transcribed in IPA as /y/ and described phonetically as a high front tense rounded

vowel. English does not have a vowel with those characteristics. It has a high front tense unrounded vowel /i/ and a high back tense rounded vowel /u/, so /y/ falls in between two English sounds. In the Perceptual Assimilation Model, comparing /y/ to either /i/ or /u/ would be considered a hard contrast to make, as it falls between two perceptual categories in English. If we apply the PAM model not from L1 to L2, but from L2 to L1, for the English-dominant bilinguals, comparing /y/ to /i/ or /y/ to /u/ would be a difficult "single category" comparison, as they are in terms of articulatory gestures (a key feature of PAM – see Werker, 2018) single category comparisons, where /y/ can be perceived as a somewhat deviant /i/ or a somewhat deviant /u/ in production. The PAM predicts that "single category" categorizations are the most difficult for people to perceive. It is possible that English-dominant bilinguals would also struggle on vowel disambiguated trials (in addition to tone trials) when comparing a vowel like /y/ with a vowel like /i/ or /u/.

Difficult vowel comparisons like /y/ to /u/ or /y/ to /i/ were not included as part of Experiment 1. A constraint of the quadruplet sets in this experiment and the original research (Quam and Creel 2017a) is that all of the stimuli words needed to be real words that met the vowel-tone quadruplet construction that could be conveyed well in a picture. As a result, most of the vowels used in the stimuli were either diphthongs or could be perceived as mapping cleanly onto English vowel sounds. Thus, it is possible that the observed effect was not limited to tone, but the experiment simply did not test vowels that were unique to Mandarin, like /y/, that would have shown a similar attrition effect. To that end, Quam and Creel (2017a) offered two mutually exclusive explanations for their findings, replicated in Experiment 1: Explanation one is that there

is something unique about tone that makes it more prone to attrition effects. Explanation 2 for the tone attrition is that consistent use of English vowel categories reinforced the particular Mandarin vowel categories that were tested in this experiment and the original research paper, masking a general attrition effect for Mandarin. These hypotheses served as the foundation for Experiment 2, currently ongoing at the Child Language Learning Center at Portland State University's Speech and Hearing Sciences Department.

Experiment 2

Methods

Participants.

To date, 21 participants have been recruited (6 male, $M = 29.35$ years, $SD = 8.15$) who self-identify as "fluent" or "proficient" native Mandarin speakers. One participant has been excluded for having more exposure to a non-Mandarin tonal dialect than to Mandarin during childhood. Primary recruitment channels were the Portland State University Office of International Students, the local Confucius Institute, local community groups and related social media groups, bulletin boards located around Portland State University's campus, and word of mouth.

Language Dominance

After completing the main familiar word experiment, participants were given the Multilingual Naming Test (MINT) (Gollan et al., 2012) and the Bilingual Dominance Survey (BDS) (Dunn & Fox Tree, 2009) described in Experiment 1. In Experiment 1, the

Mandarin MINT range was 18-64 (out of 68), with an average of 48.24, and a standard deviation of 12.09. In Experiment 2, the Mandarin MINT range is 61-68 (out of 68), with an average of 64, and standard deviation of 2.26. This means that our current sample is highly Mandarin dominant, particularly compared with the sample from Experiment 1. We believe that as we modify our recruitment efforts away from the International Student Office and local Confucius Institute towards local Chinese-American groups in the Portland Area, we will be able to recruit a wider variety of language backgrounds, including more English-dominant bilinguals. This experiment has only tested twenty people, so the sample characteristics are expected to converge on the sample population characteristics from Experiment 1 as we continue to recruit from the broader Mandarin-speaking population of Portland. Table 3 presents a summary of language dominance profiles for this experiment compared to Experiment 1 and to Quam and Creel (2017a).

Table 3: Participant language dominance measures across experiments. This table shows how language dominance profiles for Experiment 2 participants compare to participants in Experiment 1 and Quam and Creel (2017a). Experiment 2 had the most Mandarin-dominant participants. Experiment 1 participants were the most English dominant. Participants in the original research paper were between the two, but closer to Experiment 1.

	Experiment 2		Experiment 1		Quam and Creel (2017a)	
Test Name	Mean *(SD)*	Range	Mean *(SD)*	Range	Mean *(SD)*	Range
Multilingual Naming Test						
Mandarin	64.05 *(2.258)*	61-68	48.24 *(12.09)*	18-67	51.47 *(10.56)*	20-64
English	53.45 *(6.878)*	37-67	59.91 *(6.78)*	33-67	57.81 *(5.87)*	40-68
Composite	10.65 *(7.028)*	0-26	-11.67 *(16.36)*	-47-27	-6.33 *(14.29)*	-38-18
Bilingual Dominance Scale	18.1 *(9.318)*	1-36	-5.34 *(11.59)*	-28-16	-.147 *(10.51)*	-23-16
Age of Arrival (years)	24.175 *(8.129)*	4-39.5	7.59 *(7.11)*	0-19	8.32 *(6.01)*	0-20

In Experiment 1 and Quam and Creel (2017a), the measures of language dominance were combined to make a single factor using principal component analysis (PCA). Unlike previous experiments, this did not yield a strongly correlated measure. The Kaiser-Meyer-Olkin measure of sampling adequacy for this factor was .595.[3] The PCA was closely correlated with almost all measures of language dominance. Correlations for the individual language measures are listed in Table 4.

Table 4: Correlations between language dominance measures. Even though the PCA is not strongly correlated to all measures as it was in Quam and Creel (2017a) and Experiment 1, it is being used for methodological consistency. It is also strongly correlated with all but one measure of language dominance.

	MINT Mandarin	MINT English	MINT Composite	BDS	AoA	PCA
MINT Mandarin		.083 (p = .727)	.140 (p = .555)	.400 (p = .08)	.378 (p = .101)	.388 (p = .091)
MINT English			-.807 (p = .001)	-.287 (p = .219)	-.159 (p = .503	-.490 (p = .028)
MINT composite				.418 (p = .066)	.327 (p = .160)	.696 (p = .001)
BDS					.614 (p =.004)	.867 (p < .001)
AoA						.823 (p < .001)

---

[3] A Kaiser-Meyer-Olkin of above .6 is desirable for adequate correlation. However, since both Experiment 1 and the original paper were conducted with a highly correlated PCA analysis, I am choosing to proceed with it, given the abnormally low variance in the Mandarin MINT condition (Mandarin MINT Variance = 5.1, English MINT Variance = 47.31), the small sample size included thus far, and how strongly the PCA is correlated to the other values of language dominance.

Experimental Design

To test if the attrition effect (shown in Quam and Creel (2017) and Experiment 1) was unique to tones, we created a list of 87 new, imageable stimuli making sure to include Mandarin vowels that do not exist in English. See Appendix B for a list of stimuli, stimuli pairs and predicted vowel-pair difficulty. The new stimuli were recorded by a Chinese-born Mandarin speaker from Henan Province on a Marantz PMD 670 recorder. Four Mandarin-speaking lab members approved the stimuli (and their tones and vowels) as sounding 'native like'.

In a pilot experiment, these auditory stimuli were presented to a group of 10 English monolinguals who were asked to categorize each Mandarin word's vowel by selecting the English vowel the sound most resembled. For this experiment, participants were played the stimuli from the first experiment as well as the newly designed stimuli and asked to pick the best fit English vowel from the selection "boat, head, heed, hid, hood, hot, hut, who, or 'other'." These words were selected to provide the monophthong vowels in a similar phonological context. Participants clicked on a button labeled with the word they thought best represented the vowel in the word they had just heard. Figure 6 shows the results for their responses.

English Classification of Mandarin Vowels
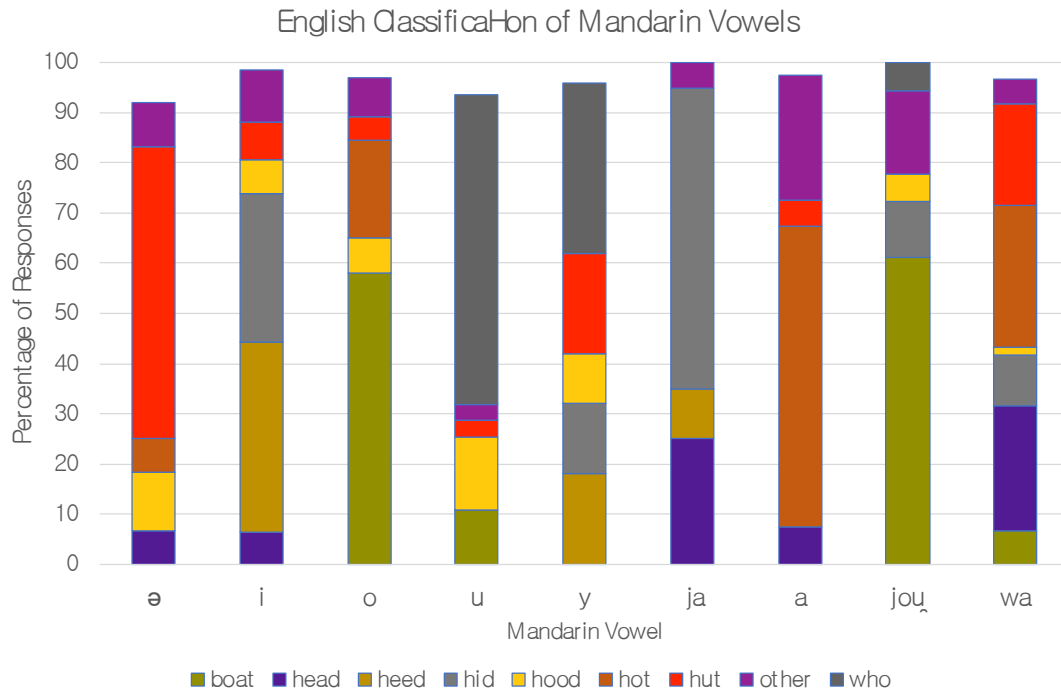
English monolinguals Mandarin stimuli vowels as best-fit English vowels. Low agreement among English monolinguals indicate which Mandarin vowels are less likely to be processed by recruiting English vowel perceptual categories.

Mandarin vowels that we had previously identified as having clear English

equivalents, what the Perceptual Assimilation Model called "categorized vowels", had a

high level of agreement. These data also show that vowels without clear English

equivalents, what the Perceptual Assimilation Model called "uncategorized" or "single

category" vowels had low levels of agreement. Together this suggests that participants

in the eye-tracked Experiment 2 will not be able to recruit English vowel categories to

process the "uncategorized" and "Single Category" vowels (See Appendix B for

predicted categories). The original predictions were that /y/ would have low agreement,

and that the mid-high back unrounded Mandarin vowel /ɤ/, another vowel that does not

exist in English, would also be hard for the English-dominant participants to

differentiate. The predictions were half right; /y/ had low agreement among our pilot

participants. Mandarin vowel /ɤ/, however, was very clearly considered to be /ʌ/ by our

participants.

The new stimuli were used to generate trials that differed on vowel only (*zhi1-zhu1*), tone only (*he1-he2*) or were designed to be maximally different (*li1-shu2)*. To enable us to include more difficult vowels, we did not constrain the words by requiring them to comprise quadruplets in which each word differed from the others in tone, vowel, or tone & vowel. Instead, we organized words into pairs that were differentiated along one of these dimensions, resulting in the trials being 'tone differentiated', 'vowel differentiated', and 'maximally different'. Some words happened to occur in more than one trial type. However, this was not required. Note that the 'maximally different' trials, which we will call 'baseline' trials, are distinct from the 'tone & vowel' trial type used in Experiment 1 and by Quam and Creel (2017a), as the words are more distinct in Experiment 2. However, this trial type was used in previous work (Quam & Creel, 2017b). These trial pairs were used to generate pseudo-randomized orders, which were checked for repeating patterns of which side of the screen the target appeared on[4], repeating patterns of trial type, and to ensure that all trial types were distributed evenly from start to end of the generated order. Each pair was presented twice with each word appearing once as target and competitor.

The auditory recordings and visual stimuli were used to build an eye-tracking experiment in Experiment Builder (SR Research, Mississauga, Ontario, Canada), modeled after the one used for Experiment 1. Participants were first exposed to the 87

---

[4] The rules for generating pseudo-randomized trials were: (1) Trial types (tone, vowel, or baseline) are intermixed and generally balanced across the list. (2) No more than three consecutive trials of any type. (3) No more than four consecutive targets on the same side. (4) Targets and competitors appear in non-repeating sequences on both sides of the screen. (5) The same word cannot appear in consecutive trials (as target or competitor).

stimuli in a familiarization phase, where the participants were looking over the course of

the trials. Participants' reaction times and accuracy were also recorded for analysis.

After completing the eye-tracking experiment, participants completed two measures of

language dominance, the Multilingual Naming Test (Gollan et al., 2012) and the

Bilingual Dominance Survey (Dunn and Fox Tree, 2009) to establish language

dominance. This study has been approved by Portland State University's Institutional

Review Board.

Data Analysis

Accuracy.

As with Experiment 1, accuracy was analyzed across all trial types. Participants

overall achieved a high level of accuracy. Average accuracy across participants was

96% across all trial types. In baseline trials, participants clicked on the correct picture

99.9% of the time ($SD = 0.3\%$). In tone-differentiated trials, participants had a mean

accuracy of 99.2% ($SD = 1.3\%$), and in vowel-differentiated trials, participants had a

mean accuracy of 99.4% ($SD = 0.9\%$) The single lowest participant accuracy by trial

type was in tone trials, at 94% accuracy. As in Experiment 1, the accuracy data was

transformed using the guidelines established as best practices in Barr (2008) (and used

by Quam and Creel (2017a), using the empirical logit (e-logit) transformation to allow for

statistical analysis. The transformed data were then used to conduct repeated

measures ANOVAs with trial type (baseline, tone, vowel) as the within-subject factor.

Trial type was a significant predictor of clicking accuracy by subject but not by item

($F_1(2,18) = 7.686$, $p = .004$; $F_2(2,58) = .88$, $p = .42$). Paired t-tests by subject revealed

that baseline accuracy was higher than tone accuracy ($t(19) = 2.536$, $p = .020$) as well

as being significantly higher than vowel accuracy (t(19) = 4.028, p = .001). There was

no significant difference in accuracy between tone- and vowel-differentiated trials (t(19)

= 1.619, p = .122). No by-item paired t-tests yielded significant results (baseline-vowel:

(t(74) = 1.418, p = .161; baseline-tone: t(80) = .650, p = .518; tone-vowel: t(60) = -.896,

p = .374).

To look for effects of language dominance, the three measures of language

dominance (MINT, BDS and AOA) were combined with Principal Component Analysis

(PCA) to create a single numerical representation of language dominance. The PCA

and measures of language dominance are discussed in the Language Dominance

section above. The PCA was used to conduct an ANCOVA. There was a significant

main effect of trial type (baseline, tone or vowel; F_1(2,17) = 8.149, p = .001) but no

significant main effect of PCA (F(1,18) = .001, p = .975) and no interaction effect of PCA

and trial type (F_1(2,17) = .051, p = .951). Figure 7 shows participant accuracy in each

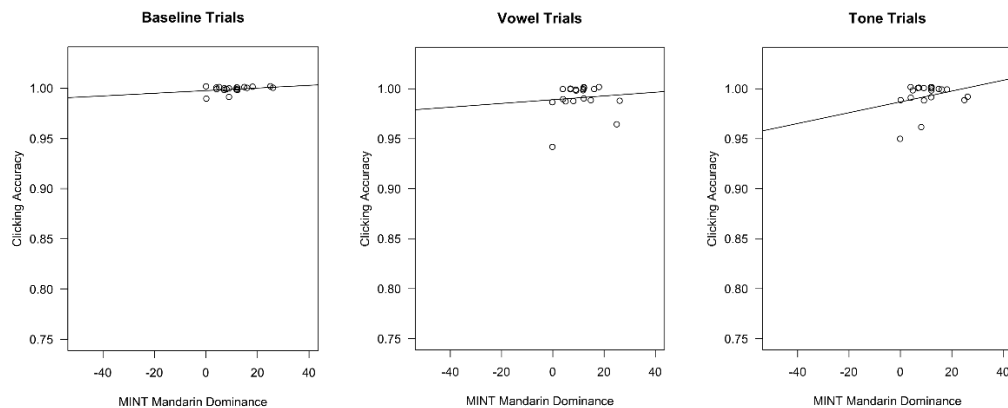trial type plotted against Mandarin dominance.

Figure 7: Participant accuracy in each trial type.

One hypothesis for the selective tone attrition effect in Experiment 1 (and Quam & Creel, 2017a) was that English-dominant bilinguals were able to recruit their English vowel categories to process Mandarin sounds. Based on the pilot study, vowel comparisons were classified by how an English-dominant speaker might apply the Perceptual Assimilation Model to the vowel comparisons. Since this experiment was designed specifically to test vowel categories, a repeated measures ANOVA by-subject was conducted with the predictor vowel difficulty level (easy, medium, or hard). This did not show a significant effect of vowel difficulty level on accuracy ($F(2) = .888$, $p = .42$).

Reaction Time

For reaction time (RT) analyses, trials that were more than three standard deviations from each participants' mean RT were excluded. This excluded approximately 3.7% of trials. ANOVAs with predictor trial type by subject and item did not meet the threshold for statistical significance for the main effect of trial type ($F_1(2) = 3.047$ $p = .059$; $F_2(2,58) = 1.731$, $p = .186$).

To determine if language dominance played a role, an ANCOVA was conducted with trial type and the PCA language dominance score as predictors. There was a significant main effect of trial type ($F(2,17) = 3.412$, p = .044). The main effect of PCA was not significant ($F(1,17) = .001$, $p = .971$). The interaction of trial type and language dominance was not significant ($F(2,17) = 2.891$, $p = .068$). Planned comparisons by trial type revealed that baseline trials ($M = .974$ sec $SD = .315$ sec) were significantly faster than vowel differentiated trials ($M = 1.01$ sec $SD = .307$ sec; $t(19) = 2.13$, $p = .046$). Baseline trials were also significantly faster than tone differentiated trials ($M = 1.05$ sec $SD = .31$ sec; $t(19) = 2.16$, $p = .044$). Tone differentiated and vowel differentiated trials

were not significantly different (t(19) = 1.20. p = .245). No correlation tests yielded

significant results (baseline r = -.098, p = .68; vowel r = -.001, p = .998; tone r = .126, p

= .598). Figure 8 shows participant RTs graphed against Mandarin-dominance.
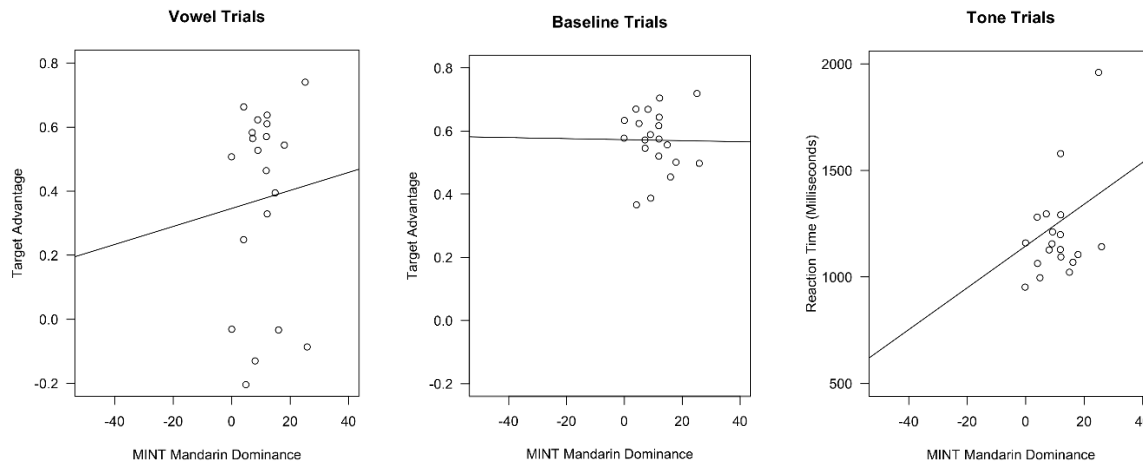
Figure 8: Reaction time versus Mandarin dominance. There was no significant correlation
between reaction time and Mandarin-dominance.

A by-subjects ANOVA was done comparing RTs across predicted difficulty of

vowel comparison (easy, medium, or hard). This test did not show a significant effect of

vowel comparison difficulty on participant's RTs (F(2,18) =1.316, p = .293).

Gaze Analysis

Participants' gaze in this experiment was recorded with the Eyelink 1000 Plus

eye-tracker. Gaze samples were used to determine where the participants were looking

from 200 milliseconds to 800 milliseconds after the trial began, because this represents

the earliest a participant is able to respond to the stimuli to when the overall asymptote

for accuracy was reached, while minimizing the number of trials extended from the end

of the actual trial so that all trials contributed to the data (11.3% of trials extended). With

this information we used custom Python scripts to calculate a participant's look to target

and competitor. Following the best practices laid out in Experiment 1, this information was put through the e-logit transformation to enable statistical analysis. This is done for target fixations and competitor fixations. The number of looks to the competitor made at the same time point were subtracted from the looks to the target, giving the target advantage. For the sake of clarity, in this paper, untransformed gaze fixation data is reported in means in the text and depicted in figures, where 0 is chance looking and 1 is looking solely at the target. An ANOVA was conducted with trial type (baseline, vowel and tone) as a within-subject predictor by-participant and by-item. This revealed that trial type was a significant predictor of looks to target ($F_1(2,18) = 41.782$, $p < .001$; $F_2(2,58) = 19.541$, $p < .001$). Planned t-tests showed that the difference between baseline ($M = .40$, $SD = .09$) and tone ($M = .27$, $SD = .08$) target advantage was significant by-subject and by-item ($t_1(19) = 5.370$, $p = .011$; $t_2(80) = 7.075$, $p < .001$). The difference between baseline and vowel ($M = .34$, $SD = .09$) differentiated trials was also significant by-subject and by-item ($t_1(19) = 9.455$, $p = .007$; $t_2(96) = 4.755$, $p < .001$). Tone-differentiated and vowel-differentiated trials were significant by-subject ($t(19) = 3.720$, $p = .002$) but not by item ($t(96) = 1.083$, $p < .282$).

A within-subjects ANCOVA across subjects with the language dominance PCA as a covariate showed a significant main effect of trial type ($F_1(2,17) = 40.360$, $p < .001$). The main effect of language dominance was not significant ($F(2,17) = .407$, $p = .533$). The interaction of target advantage and language dominance was not significant ($F(2,17) = .189$, $p = .670$). No correlation testing was done. However, Figure 9 shows target advantage plotted against Mandarin dominance in each trial type.
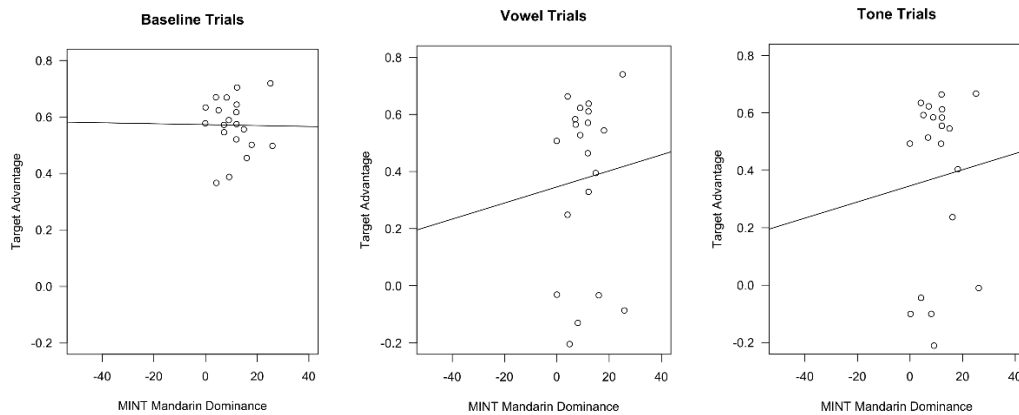
Figure 9: Correlations Between Gaze and Language Proficiency. There is no correlation between trial type and language dominance, but tone-differentiated trials were significantly slower than baseline or vowel trials when sorted by subject.

Again, since vowels were of particular significance to this study, a within-subject ANOVA was conducted for vowel-differentiated trials based on the vowel comparison categories. This revealed that vowel categorization (easy, medium, or hard) was a not-quite-significant predictor of target advantage ($F(2,18) = 3.136$, $p = .055$).

An ANCOVA with vowel categorization level as a within-subject factor with language dominance as a between-subject covariate had a not-quite-significant main effect of trial type ($F(2,17) = 2.975$, $p = .064$) and a not-quite-significant main effect of language dominance ($F(1,17) = 3.977$, $p = .061$). There was also no interaction effect on target advantage in vowel trials ($F(2,17) = .269$, $p = .61$).

Raw (untransformed) target advantage, where 0 is chance looking and 1 would signify 100% looks to target, indicates that participants looked least often to the target in the medium comparison trial. Medium difficulty comparisons had the lowest target advantage, with a target advantage of 26.61%. Participants had a target advantage of 35.63% in easy comparison trials and the hard comparison trials had a target advantage of approximately 34.45% of the time. However, it is very likely that the small sample size

and overall Mandarin dominant nature of this sample explains this counterintuitive

finding.

Discussion

Accuracy overall was very high in this study, although there were some

differences between tone-differentiated and baseline trials as well as between vowel-

differentiated trials and baseline trials. Reaction times across trial type and vowel pair

comparisons were comparable and yielded no significant differences. It is only by

analyzing target fixations (gaze data) that significant differences by vowel comparison

begin to emerge. The most unusual finding, given the results of Experiment 1 and the

results published in Quam and Creel (2017a), is that there were no discernible effects of

language dominance on looks to target. However, given the highly Mandarin-dominant

nature of our current list of participants, this is not too surprising. The findings of

different looks across trial types was robust enough to appear in analyses by-subject

and by-item. However, it is contrary to expectations given previous data that paired t-

tests revealed the significant differences are found between baseline and vowel-

differentiated trials, not between baseline and tone or between tone and vowel-

differentiated trials.

The vowel comparison analysis is also interesting, since significant differences in

vowel comparison pairs only appear in patterns of gaze fixation, not in analysis of

accuracy or reaction time. It is surprising that easy and hard vowel-contrast trials do not

exhibit significantly different sample means, and that both are significantly different from

the medium trials. However, given the small sample size, and highly dominant Mandarin

language profile of all participants, along with the low power of the experiment as is,

speculation into why is best deferred until the experiment is properly powered and we

see if this effect remains.

The results of this study raise some interesting questions. Many of the findings

from the new experiment seem to contradict the findings from the previous study.

However, an important caveat must be mentioned regarding the new study's data: the

small sample size. Since this study only includes data from 20 participants to date, the

power of this experiment is likely too low to detect the predicted effects. A power

analysis conducted in R indicated that for 80% power, at $\alpha = .05$, depending on which

metrics were being compared, the sample size needs to be between 46 and 90

participants. Depending on the metrics being compared, the current power of the

experiment averages about 54%. Given that, it is safe to conclude that there is not

enough signal among all the noise.

Second, a comparison of language-dominance measures indicates that the two

sample populations are inherently unequal. Comparing the average age of arrival of the

two groups, there is a difference of approximately 14 years ($M_1 = 7.108$ years, $M_2 =$

24.175 years), a significant difference (unequal variance $t(29) = -8.11$, $p < .001$).

Examination of the BDS and MINT dominance scores also yields significant differences

(see Table 3). Participants in Experiment 2 are not just more Mandarin-dominant than

participants in Experiment 1, they are, on the whole, more Mandarin-dominant than the

most Mandarin-dominant participants of Experiment 1 and in Quam and Creel (2017a).

This means they would not be expected to show any attrition effects in tone processing,

so the data patterns exhibited in this subsample are expected.

Conclusion

Experiment 1 largely replicated the findings in Quam and Creel (2017a),

confirming the findings in that paper. Quam and Creel (2017a) offered two explanations

for the finding of tone processing attrition which was confirmed in Experiment 1. The

first explanation was that the tone attrition was due to the unique way the brain stores

and accesses lexical tone. The other explanation is that English-dominant bilinguals

were able to recruit English vowel perceptual categories to process Mandarin sounds.

Experiment 2 was designed to differentiate these two hypotheses by directly testing the

second hypothesis. Although it is underpowered at the moment, the experimental

paradigm is sound and once properly powered should provide a thoughtful test to the

question of whether the tone attrition pattern detected was attributable to the unique

characteristics of how the brain stores tones or a more general attrition effect masked in

the English-dominant bilingual vowel scores because of their ability to recruit English

vowel perceptual categories.

This line of research is foundational to understanding how the phonologies of

different languages interact in an individual. Together, Quam and Creel (2017a),

Experiment 1, and Experiment 2 examine how a second language changes a speaker's

ability to perceive and identify sounds in their first language. Knowledge of how a

second language impacts a speaker's first language has potential impacts on how to

best support multilingual speakers to effectively interact with all their communication

partners regardless of the language context. This foundational knowledge will help to

correct the historical trends (and continuing issues) of over-identification and under-

identification of bilingual individuals as having communication disorders (Crawley,

2018).

As the number of people who are required to be functional bilinguals increases,

more bilinguals will require support from communication specialists faced with the

question "is this communication deviation a difference arising from language proficiency

or is this a communication disorder that requires intervention for a speech disorder and

support to communicate effectively?" It is only when clinicians, educators and

therapists have access to a more complete idea of how languages interact with each

other in their clients that they will be able to best support the 60 million Americans who

speak a language other than English at home.

Appendix A: Experiment 1 Stimuli

Stimuli in this appendix were taken directly from Quam and Creel (2017a). Stimuli quadruplets were designed to have the same onset, and differ in vowel, tone or tone and vowel. A further restriction was that all stimuli needed to be easily conveyed in picture form.

1

cha1   cha2

chuang 1   chuang2

2

mu3   mu4

mi3   mi4

3

shi1   shi3

shu1   shu3

4

xian1   xian4

zin1   xin4

5

hua1   hua4

he 1   hua4

6

qiu1   qiu2

quan1   quan2

7

ta3   ta4

tu3   tu4

Appendix B: Experiment 2 Stimuli

Stimuli are presented by trial type: Baseline, Vowel and Tone.

1 - Baseline

Baseline trials have different onset, so no difficulty comparisons were done.

Word 1	Word 2	Word 1	Word 2
ben1	ling2	jün1	zheng3
ci1	lu4	ke3	yün4
cu1	meng2	ke4	hong2
cu4	mi3	ku1	zhong3
deng1	mi4	ku3	zhong4
deng4	mu3	li4	zhu3
di2	mu4	bin1	zhu2
ding1	she2	deng3	xin1
ding3	song1	ding4	tong1
dong4	shi3	dong1	ting2
feng1	shu3	gong1	hen2
feng2	shu4	ting3	zhong1
feng4	nü3	hen4	ming2
ge1	zhi3	ni3	lü4
ge2	zhi2	ci4	zhun3
gu1	tong2	di3	shu1
gu4	tong3	du2	geng1
he1	tu3	du3	she4
he2	tu4	xin4	lü2
he4	yin2	ku4	ling3
heng2	yin4	li2	shu1
hong1	yün2	shi1	zhen3
hu2	song4	lu2	hen4
hun2	ting1	zhu1	ming2
jin1	shu2	gu3	song1

2 - Vowel

Vowel-differentiated comparisons, with expected level of difficulty (ELD). Expected level of difficulty is based on vowel distinctive features and PAM predicted categorizations. Based on predictions, we predicted that comparing pinyin vowels "e" and "o" (IPA transcriptions: /ɤ/ and /o/), as well as pinyin vowel "ü" to "i" or "ü" to "u" (IPA transcriptions: /y/ to /i/ and /y/ to /u/) would be hard for English-dominant participants to distinguish, as both were predicted to be mapped onto similar acoustic space in the English vowel perceptual categories. Easy comparisons are where both vowels clearly mapped onto distinctive areas of the phonemic inventory. Medium vowel comparisons are more articulatory and perceptively unique than easy vowel comparisons, but not has similar as the hard vowel comparisons.

Word 1	Word 2	Vowel contrast	ELD	Word 1	Word 2	Vowel contrast	ELD
ben1	bin1	e,i	E	lu2	lü2	u,ü	H
ci1	cu1	i,u	E	lu4	lü4	u,ü	H
ci4	cu4	i,u	E	meng2	ming2	e,i	E
deng1	ding1	e,i	E	mu3	mi3	u,i	E
deng1	dong1	e,o	H	mu4	mi4	u,i	E
deng3	ding3	e,i	E	ni3	nü3	i,ü	H
deng4	ding4	e,i	E	she2	shu2	e,u	M
deng4	dong4	e,o	H	she4	shu4	e,u	M
di2	du2	i,u	E	shi1	shu1	i,u	E
di3	du3	i,u	E	shi3	shu3	i,u	E
ding1	dong1	i,o	E	ting1	tong1	i,o	E
ding4	dong4	i,o	E	ting2	tong2	i,o	E
ge1	gu1	e,u	M	ting3	tong3	i,o	E
geng1	gong1	e,o	H	yün2	yin2	ü,i	H
he2	hu2	e,u	M	yün4	yin4	ü,i	H
hen2	hun2	e,u	M	zhen3	zhun3	e,u	M
hong2	heng2	o,e	H	zheng3	zhong3	e,o	H
jün1	jin1	ü,i	H	zheng4	zhong4	e,o	H
ke3	ku3	e,u	M	zhu2	zhi2	u,i	E
ke4	ku4	e,u	M	zhu3	zhi3	u,i	E
li2	lu2	i,u	E				

3 – Tone

Tone-differentiated pairs of stimuli, with expected level of difficulty (ELD). ELD is determined by similarity of pitch contours.

Word 1	Word 2	Tone contrast	ELD
ci1	ci4	1,4	M
cu1	cu4	1,4	M
deng1	deng3	1,3	E
deng1	deng4	1,4	M
deng3	deng4	3,4	M
di2	di3	2,3	H
ding1	ding3	1,3	E
ding4	ding3	4,3	M
dong1	dong4	1,4	M
du2	du3	2,3	H
feng1	feng2	1,2	M
feng1	feng4	1,4	M
feng4	feng2	4,2	E
ge1	ge2	1,2	M
gu1	gu4	1,4	M
gu3	gu4	3,4	M
he1	he2	1,2	M
he1	he4	1,4	M
he2	he4	2,4	E
hen2	hen4	2,4	E
hong1	hong2	1,2	M
ke4	ke3	4,3	M
ku1	ku3	1,3	E
ku4	ku3	4,3	M
li4	li2	4,2	E
ling2	ling3	2,3	H

Word 1	Word 2	Tone contrast	ELD
lu4	lu2	4,2	E
lü4	lü2	4,2	E
mi3	mi4	3,4	M
mu3	mu4	3,4	M
she2	she4	2,4	E
shi1	shi3	1,3	E
shu1	shu3	1,3	E
song1	song4	1,4	M
ting1	ting2	1,2	M
ting3	ting2	3,2	H
tong1	tong2	1,2	M
tong3	tong2	3,2	H
tu3	tu4	3,4	M
xin1	xin4	1,4	M
yin2	yin4	2,4	E
yün2	yün4	2,4	E
zheng3	zheng4	3,4	M
zhi2	zhi3	2,3	H
zhong1	zhong3	1,3	E
zhong1	zhong4	1,4	M
zhong3	zhong4	3,4	M
zhu1	zhu3	1,3	E
zhu1	zhu2	1,2	M
zhu2	zhu3	2,3	H

Works Cited

Barr, D. J. (2008). Analyzing 'visual world' eyetracking data using multilevel logistic regression. *Journal of Memory and Language*, *59*(4), 457–474. https://doi.org/10.1016/j.jml.2007.09.002

Best, C. T., Goldstein, L., Tyler, M. D., & Nam, H. (2009). Articulating the Perceptual Assimilation Model (PAM): Perceptual assimilation in relation to articulatory organs and their constriction gestures. *The Journal of the Acoustical Society of America*, *125*(4), 2758–2758. https://doi.org/10.1121/1.4784648

Best, C. T., McRoberts, G. W., & Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *The Journal of the Acoustical Society of America*, *109*(2), 775–794. https://doi.org/10.1121/1.1332378

Bosch, L., & Sebastián-Gallés, N. (2003). Simultaneous Bilingualism and the Perception of a Language-Specific Vowel Contrast in the First Year of Life. *Language and Speech*, *46*(2–3), 217–243. https://doi.org/10.1177/00238309030460020801

Burnham, D., Kim, J., Davis, C., Ciocca, V., Schoknecht, C., Kasisopa, B., & Luksaneeyanawin, S. (2011). Are tones phones? *Journal of Experimental Child Psychology*, *108*(4), 693–712. https://doi.org/10.1016/j.jecp.2010.07.008

Chang, C. B. (2013). A novelty effect in phonetic drift of the native language. *Journal of Phonetics*, *41*(6), 520–533. https://doi.org/10.1016/j.wocn.2013.09.006

Cheung, H., Chen, H.-C., Lai, C. Y., Wong, O. C., & Hills, M. (2002). The Development of Phonological Awareness: Effects of Spoken Language Experience and Orthography. *Cognition*, *81*(3), 227–41. https://doi.org/10.1016/S0010-0277(01)00136-6

Crawley, C. J. (2018, October). *Fundamentals of Culturally and Linguistically Appropriate Evaluations*. Presented at the Oregon Speech Language and HEaring Association, Salem, OR.

Da, J. (2004). A corpus-based study of character and bigram frequencies in Chinese e-texts and its implications for Chinese language instruction. In Z. Pu, T. Xie, & J. Xu (Eds.), *The studies on the theory and methodology of the digitalized Chinese Conference on New Technologies in Teaching and Learning Chinese* (pp. 501-511). Beijing, China: Tsinghua University Press.

Da, J. (2005). *Modern Chinese character frequency list.* Retrieved from http://lingua.mtsu.edu/chinese-computing/statistics/char/list/php?Which=MO

Dunn, A. L., & Fox Tree, J. E. (2009). A quick, gradient Bilingual Dominance Scale*. *Bilingualism: Language and Cognition*, *12*(03), 273. https://doi.org/10.1017/S1366728909990113

Gildersleeve-Neumann, C. E., Peña, E. D., Davis, B. L., & Kester, E. S. (2009). Effects on L1 during early acquisition of L2: Speech changes in Spanish at first English contact*. *Bilingualism: Language and Cognition*, *12*(2), 259–272. https://doi.org/10.1017/S1366728908003994

Gollan, T. H., Weissberger, G. H., Runnqvist, E., Montoya, R. I., & Cera, C. M. (2012). Self-ratings of spoken language dominance: A Multilingual Naming Test (MINT) and preliminary norms for young and aging Spanish–English bilinguals. *Bilingualism: Language and Cognition*, *15*(03), 594–615. https://doi.org/10.1017/S1366728911000332

Grosjean, F. (1989). Neurolinguists, beware! The bilingual is not two monolinguals in one person. *Brain and Language*, *36*(1), 3–15. https://doi.org/10.1016/0093-934X(89)90048-5

Kaiser, H. (1974). An Index of Factorial Simplicity. *Psychometrika*(39), 31–36.

Kohnert, K. (2013). *Language disorders in bilingual children and adults* (2nd ed.). San Diego, California ; Oxford, [England] ; Melbourne, [Australia]: Plural Publishing.

Malins, J. G., & Joanisse, M. F. (2010). The roles of tonal and segmental information in Mandarin spoken word recognition: An eyetracking study. *Journal of Memory and Language*, *62*(4), 407–420. https://doi.org/10.1016/j.jml.2010.02.004

Polka, L., & Werker, J. F. (1994). Developmental changes in perception of nonnative vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance*, *20*(2), 421–435. https://doi.org/10.1037/0096-1523.20.2.421

Quam, C., & Creel, S. C. (2017a). Mandarin-English Bilinguals Process Lexical Tones in Newly Learned Words in Accordance with the Language Context. *PLoS ONE*, *12*(1), e0169001. https://doi.org/10.1371/journal.pone.0169001

Quam, C., & Creel, S. C. (2017b). Tone Attrition in Mandarin Speakers of Varying English Proficiency. *Journal of Speech, Language & Hearing Research*, *60*(2), 293–305. https://doi.org/10.1044/2016_JSLHR-S-15-0248

Reid, A., Burnham, D., Kasisopa, B., Reilly, R., Attina, V., Rattanasone, N. X., & Best, C. T. (2015). Perceptual assimilation of lexical tone: The roles of language experience and visual information. *Attention, Perception, & Psychophysics*, *77*(2), 571–591. https://doi.org/10.3758/s13414-014-0791-3

Schirmer, A., Tang, S.-L., Penney, T. B., Gunter, T. C., & Chen, H.-C. (2005). Brain Responses to Segmentally and Tonally Induced Semantic Violations in Cantonese. *Journal of Cognitive Neuroscience*, *17*(1), 1–12. https://doi.org/10.1162/0898929052880057

Taft, M., & Chen, H.-C. (1992). Judging Homophony in Chinese: The Influence of Tones. In H.-C. Chen & O. J. L. Tzeng (Eds.), *Advances in Psychology* (pp. 151–172). https://doi.org/10.1016/S0166-4115(08)61891-9

Tyler, M. D., Nil, M. D., Nil, M. D., Best, C. T., Faber, A., & Levitt, A. G. (2014). Perceptual Assimilation and Discrimination of Non-Native Vowel Contrasts. *Phonetica*, *71*(1), 4–21. https://doi.org/10.1159/000356237

US Census Bureau. (n.d.). Detailed Languages Spoken at Home and Ability to Speak English. Retrieved February 27, 2019, from https://www.census.gov/data/tables/2013/demo/2009-2013-lang-tables.html

Wang, W., & Li, K. (1967). Tone 3 in Pekinese. *Journal of Speech and Hearing Research, 10*(3), 629-36.

Werker, J. F. (2018). *Perceptual beginnings to language acquisition. 39*(4), 703–728. https://doi.org/10.1017/S0142716418000152

Ye, Y., & Connine, C. M. (1999). Processing Spoken Chinese: The Role of Tone Information. *Language and Cognitive Processes*, *14*(5–6), 609–630. https://doi.org/10.1080/016909699386202