11-26-1980

An Evaluation of a Talking Machine: the HC 120 Phonic Mirror Handivoice

James Joseph Oggerino
Portland State University

AN ABSTRACT OF THE THESIS OF James Joseph Oggerino for the Master of Science in Speech Communication presented November 26, 1980.

Title: An Evaluation of a Talking Machine: The HC 120 Phonic Mirror Handivoice.

APPROVED BY MEMBERS OF THE THESIS COMMITTEE:

[redacted]

Steve Kosokoff, Chairman

[redacted]

Theodore Grove

[redacted]

Robert H. English

This study sought to determine the basic intelligibility of synthesized speech as produced by the HC 120 Phonic Mirror Handivoice. It involved 48 male and female subjects divided into two groups. Group 1 had 40 subjects ranging in age from 18 to 44 years. Group 2 had eight subjects ranging in age from 18 to 33 years. All subjects were screened for normal hearing. Testing for normal hearing was done in an audiological suite at Portland State University. The study was designed to determine the degree to which a group of normal hearing college students could recognize words and phrases as produced by the HC 120 Phonic Mirror Handivoice.

In order to test the intelligibility of words and phrases produced by the HC 120 Phonic Mirror Handivoice, two 20-item tests were constructed. Test Number 1 was comprised of eight spondee words, eight phonetically balanced (PB) monosyllabic words, and four phrases taken from the HC 120 brochure for a total of 20 scorable items. Test Number 2 contained eight bisyllabic words, eight monosyllabic words contained in the HC 120 brochure, and the same four phrases mentioned in Test Number 1. Both tests, therefore, contained 20 scorable items.

Results indicated that with the exception of hungry, all words (ham, headlight, his, hello) beginning with the consonant /h/ were difficult to discriminate. Two spondee words, oatmeal and birthday, were very poorly discriminated, as were the PB words tie and chair. Monosyllabic words were evenly divided between those discriminated and those not discriminated. The initial consonants of like and bike were not discriminated and neither were the words. Bread was consistently identified as thread, while fine was not discriminated at all.

Phrase identification for both Group 1 and Group 2, containing more contextual cues, was very well discriminated. It would seem then, the HC 120 had limited capabilities in the production of isolated words. As far as words and phrases organic to itself, the HC 120 had good to excellent capabilities.

AN EVALUATION OF A TALKING MACHINE: THE HC 120 PHONIC MIRROR HANDIVOICE

by

JAMES JOSEPH OGGERINO

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

MASTER OF SCIENCE
in
SPEECH COMMUNICATION

Portland State University
1980

TO THE OFFICE OF GRADUATE STUDIES AND RESEARCH:

The members of the Committee approve the thesis of James Joseph Oggerino presented November 26, 1980.

[redacted]

Steve Kosokoff, Chairman

[redacted]

Theodore Grove

[redacted]

Robert H. English

APPROVED:

[redacted]

Frank Roberts, Acting Head, Department of Speech Communication

[redacted]

Stanley E. Rauch, Dean, Graduate Studies and Research

ACKNOWLEDGMENTS

My sincere appreciation goes to Steve Kosokoff, my committee chairman and graduate advisor, for support and time generously given. My special thanks to my committee members Theodore Grove for his assistance with statistics and perseverance and to Robert H. English for his many constructive critiques.

A special thanks is due to HC Electronics (Mill Valley, California) for their generous loan of the HC 120 Phonic Mirror Handivoice.

I dedicate this thesis to my wife, Alice Oggerino, for her support and particularly her monumental patience during my graduate study. Without her unstinting help this paper would not have been possible.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER I

INTRODUCTION AND STATEMENT OF THE PROBLEM

INTRODUCTION

From early times man has been interested in talking machines. His interest, of course, does not date back to the dawn of history when man began to use sounds instead of hand signals as a means of communication (Robinson, 1979). Rather, it had its beginning in the Greek and Roman civilizations when oracles "spoke" to their supplicants by means of hidden communication tubes (Flanagan, 1976).

As far as is known, however, the first serious attempt to simulate the human voice by mechanical means was in the colossal statue of Memnon at Thebes. This was built in the eighteenth Egyptian dynasty about 1490 B.C. Memnon was supposed to emit a vocal greeting to his mother each morning at sunrise. Strabo, who visited the statue in 7 A.D., testified it did produce some sort of sound (Read and Welch, 1959).

After lying dormant for some two millenia, the interest in talking machines was revived somewhere around 1742 by two Frenchmen who built automatons. Vaucanson built an android which played a flute and a duck that drank water, ate corn, and swallowed it with a complete simulation of the digestive process. Le Droz built a writing child, while his son made a bullfinch that would jump up from a snuffbox, pour forth a melodious song, then dart back into the box as the lid was

closing (Dudley and Tarnoczy, 1950).

In 1796, Wolfgang von Kempelen built the first speaking machine of consequence. Using a drone reed from a bagpipe, von Kempelen obtained some good vowel distinctions, forming a fair /a, o, and u/. He also obtained a fair consonant sound for /p, m, and l/. His device produced semi-vowels, stops, fricatives, the transitionals h, w, and the German j. Additionally, it produced the German ch for a total of 19 consonant sounds (Dudley and Tarnoczy, 1950; Moses, 1964).

In the century preceding von Kempelen, the Imperial Academy of St. Petersburg (Moscow) offered a prize to anyone explaining the physiological differences in producing the five vowels /a, e, i, o, and u/. A German, Christian Gottlieb Kratzenstein, won the prize in 1779 by constructing five tubes which roughly approximated the size and shape of the vocal passage. All were energized by free reeds except the i tube, which was blown into directly (Dudley and Tarnoczy, 1950).

The culmination of automaton building seems to have been reached in 1860, when Herr Joseph Faber of Vienna built an intricate talking machine, which, according to Read and Welch (1959), had a tube attached to its nose when it spoke French.

The age of electronics has afforded scientists the opportunity to improve yesterday's accomplishments. One such improvement was the VOCODER (derived from VOice CODER), which was originally conceived as a means of transmission of the speech signal. The VOCODER was the brain child of Homer Dudley of the Bell Telephone Laboratories, who had sketched the device in his technical notebook in October, 1928 (Schroeder, 1966). VOCODER was largely a generic term applied to transmission systems on analysis and synthesis of the speech signal

(Encyclopedia of Science and Technology, p. 398).

An addition was added to the VOCODER and the resulting combination was exhibited at the World's Fair in New York City and San Francisco in 1939. Designed in the Bell Telephone Laboratories as a scientific novelty, the device was called the VODER. The name comes from the key letters of Voice Operation DemonstratER. The VODER was the first machine featuring the electrical operation of the human voice that was displayed publicly (Anonymous, 1939).

In the field of speech pathology today, there is available the HC 120 Phonic Mirror Handivoice, Model II, which, according to the manufacturers HC Electronics, actually talks. The brochure (1979) states the instrument is a hand-held electronic voice synthesizer which can produce virtually any word in the English language. The HC 120, the instrument with which we will be concerned, operates through a 3-digit numerical coding, and has a 16-button keyboard. Pre-programmed with 893 words, it also emits 16 short phrases (e.g., "I want. . . ."), and features a relatively unlimited vocabulary by programming combinations of morphemes, phonemes, and words. It has a memory bank and can spell words, one letter at a time, if necessary. Additionally, it contains auxiliary breath, muscle, or hand switch controls to accommodate the severely impaired.

STATEMENT OF THE PROBLEM

The present investigation sought to determine the basic intelligibility of synthesized speech as produced by the HC 120 Phonic Mirror Handivoice. Specifically, the investigation sought to answer the question: Can a group of normal-hearing adults recognize words and

phrases produced by the HC 120 Phonic Mirror Handivoice.

CHAPTER II

REVIEW OF THE LITERATURE

Prior to the advent of electricity, non-oral and oral tests of sound and speech discrimination were simple. They consisted of the watch tick and the coin-click, both high frequency sounds of around 2,000 Hz. Conversational voice and whisper tests also were used. The latter two covered the frequency ranges important to understanding speech, 400 to 3,000 Hz (Fletcher, 1950; Davis, 1970; Newby, 1972).

In the watch tick test a physician would hold his watch next to the patient's ear and instruct him to signal when he no longer heard the tick. The physician then would move the watch away from the ear and approximate the distance and degree of hearing loss, if any. Due to the absence of noisy watches, this test would seem to be impractical today. The coin-click test consisted of dropping a large coin on a hard surface. If the patient heard a "thud," he was presumed to have a high frequency hearing loss. If he heard the coin "ring," his hearing acuity was considered intact (Davis, 1970; Newby, 1972).

Conversational voice tests were conducted by placing the patient a prescribed distance from the examiner. The patient was instructed to repeat the numbers or words he heard. The distance the examiner had to move toward the client was an indication of the degree of hearing loss (Davis, 1970; Newby, 1972). Fletcher (1950) writes ". . . the maximum distance the normal ear can interpret called numbers is 40 feet. . . ." According to Davis (1970) and Newby (1972), however, the "normal-

hearing" person understands the whispered voice at 20 feet.

Because they emitted pure tones, tuning forks of various frequencies based on the "C" octaves of the scientific scale also were used to measure a person's hearing sensitivity. The scale used, octave by octave, included 64, 128, 256, 512 . . . 8,192 Hz. The vibrating tuning fork was held next to the patient's ear. The length of time the sound emitted by the tuning fork was heard determined the individual's hearing sensitivity (Davis, 1970; Newby, 1972). Newby (1972) writes: ". . . the most common fork tests are the Rinne, Weber, Bing and Schwabach, named after their nineteenth century German originators." All four methods tested the patient's hearing by bone conduction. Schwabach's method was considered quantitative, the others qualitative (Davis, 1970; Newby, 1972).

Glorig (1965) reports that tuning forks, as used by Hartman (1878), Hughes (1879), and Dean and Bunch (1919), culminated in 1922 when Fowler and Wegel developed the first commercially produced audiometer, the Western Electric 1A. A later model, the Western Electric 4A (now 4C), was constructed along different lines. It produced spoken numbers instead of pure tones (Bunch, 1947; Hudgins, Hawkins, Karlin, and Stevens, 1947).

In the 4A digits were spoken in groups of three. In the 4C digits were spoken in pairs. The 4C, however, was not a precision instrument, and its 33 dB range limited its usefulness (Hudgins et al., 1947).

The use of word stimuli rather than pure tones prompted other oral testing methods. Possibly the first of these was the round-robin method used by Jones (1934) to measure speaker intelligibility.

Students were tested in groups of 11. Ten members served as listeners as each student read ". . . one of 20 word lists . . . the listeners sat at right angles to the speaker, 30 feet removed (Black, 1957)." Scores were based on the proportion of ". . . correct responses among the 200 responses to the list of each speaker (Black, 1957)."

World War II brought with it the need to develop testing methods for the evaluation of military communications equipment. Some of the tests derived from these investigations were developed and standardized at the Psycho-Acoustic Laboratory (PAL) of Harvard University (Hirsh, Silverman, Reynolds, Eldert, and Benson, 1952; Davis, 1970; Newby, 1972; Denes and Pinson, 1973).

PAL Auditory Test Number 9 consisted of familiar two-syllable words pronounced with equal stress on both syllables. These were referred to as spondee words (Davis, 1970; Newby, 1972).

Hirsh and associates (1952) modified and improved the PAL spondee lists by constructing a test identified as the Central Institute for the Deaf (CID) Auditory Test Number 1.[1] This was achieved by combining the 84 spondee words in PAL Auditory Tests Number 9 and Number 14 and assigning judges the task of rating the words on a three-point scale of familiarity. The result was the selection of a single list of the 36 most familiar bisyllabic words. Six scramblings were made of this list. It differs from the PAL lists in that ". . . the vocabulary is confined to very familiar words suitable for children as well as adults (Benson, Davis, Harrison, Hirsh, Reynolds, and Silverman, 1951).

The Harvard Psycho-Acoustic Laboratory also constructed 24 lists

---

[1] W-1 means Word List Number 1 (Benson et al., 1951).

of 50 phonetically balanced (PB) monosyllabic words known as the PB-50 lists (Egan, 1948). According to Davis (1970), phonetically balanced means ". . . nearly all the phonemes of the English language are represented in every list of 50 words."

The PB lists were reduced to 20 by Egan (1948), who sought to improve phonetic balance, structure, and range of difficulty, while retaining a sample that is representative of English speech.

Lehiste and Peterson (1959) considered the phonetic balance of the Harvard PB-50 lists inadequate and developed a new monosyllabic word test (Goetzinger, 1972). The test consisted of 500 monosyllabic words of the consonant-word-nucleus-consonant or CNC type. These were selected from 1,263 monosyllabic words contained in the Thorndike and Lorge list of one million words, where they had appeared at least once. The new word lists were comprised of new words and old ones from the PB lists in approximately equal numbers.

The PB lists published by Egan (1948) were considered too large for many clinical patients by Hirsh et al. (1952). A more rigid application of phonetic balance and familiarity resulted in a smaller test vocabulary, which became known as CID Auditory Test W-22 or, more simply, as CID W-22. This test consisted of ". . . a vocabulary of 200 monosyllabic words divided into four groups of 50 words each. Each list was phonetically balanced (Hirsh et al., 1952)." One hundred twenty of the words were chosen from a pool of PB-50 words after five judges rated the words for familiarity. One reason for using 50 words was the ease of converting responses into percentage scores (Elpern, 1961).

Dissatisfied that the findings of Hirsh et al. (1952) showed no

difference in the degree of difficulty among the words in the W-22 word lists, Elpern (1960) gathered information from audiological clinics attached to Veterans Administration Hospitals in six large American cities. A pool of 1,490 monaural discrimination scores were accumulated and analyzed.

Elpern (1960) found differences in both average level and average range of difficulty among the four W-22 lists. Further, he suggested that if bias resulting from differences in difficulty were to be avoided, investigators should use combinations of Lists 2 and 3, and Lists 3 and 4 as speech stimuli in laboratory studies.

Giolas and Epstein (1963) compared intelligibility scores on the Harvard PB-50 word lists, CID W-22 word lists, and a 15-minute sample of continuous discourse, which was defined as ". . . representative speech in everyday situations. Subjects were 175 normal-hearing college students enrolled in general speech classes, who were tested both on the word lists and on the information presented in normal discourse.

Higher scores were obtained with the W-22 word lists than with the PB-50 lists. From this finding they concluded word lists have diagnostic value. No accurate intelligibility score was obtained on continuing discourse; hence, their prognostic value was limited.

Pickett and Pollack (1963) investigated the intelligibility of words removed from tape-recorded readings of a prescribed text by four experienced talkers. Three rates of speed were used: "deliberately slow," "normal," and "deliberately fast." The resulting range was 3.0 to 4.0 syllables per second (syl/sec) for slow utterances, 4.4 to 5.5 syl/sec for normal utterances, and 6.2 to 7.7 syl/sec for fast utter-

ances. Sections containing one to seven words were extracted from the tapes and played back to 15 college students who wrote down what they heard. The study indicated intelligibility increased as the duration of the speech sample increased (Table I).

TABLE I

FIGURES BELOW SHOW RESULTS OF A WORD INTELLIGIBILITY STUDY BY PICKETT AND POLLACK (1963)*

	Time in Seconds			% Words Correct		
	Fast	Nl	Slow	Fast	Nl	Slow
1 Word	.17	.23	.31	41	55	68
2 Words	.33	.44	.63	57	72	88
3 Words	.54	.72	1.08	76	84	98

*Percentage of words correct was transcribed from a word intelligibility diagram using geometric figures and was averaged as closely as possible by the current investigator. Time in seconds was faithfully reproduced.

In a related study Pollack and Pickett (1963) examined the intelligibility of conversational speech by excising words recorded from a fluent stream of speech at fast and slow rates. Four female college students were used as talkers. One to 15 successive words were recorded on a test tape, which was played back on a high quality sound system. Average rates of syl/sec were 6.0 at slow speed, 7.58 syl/sec at the fast rate.

Twenty-two to 29 listeners taken from a pool of 30 college students were used for each test, which consisted of 89 to 114 samples for

each talker. Listeners were informed of the number of words in each sample.

The conclusions reached by Pollack and Pickett (1963) were much the same as in their previous study: The average intelligibility of the excised samples increased with the duration of the extracted utterances, and was relatively free of the average rate of speaking.

Syllable duration was one of the variables investigated by Draegert (1951) in his study of the relationship between voice variables and speech intelligibility under high-level noise. Eighty-eight college students were used to record the speech samples, which consisted of a 74-word phonetically treated prose paragraph and a 24-word multiple choice intelligibility test. The recorded lists were played to panels of 10 listeners. Mean syllable duration for the word lists was .256 seconds and for the prose .157 seconds. From his study Draegert (1951) concluded the voice variables most closely related to speaker intelligibility in high-level noise were vocal intensity and syllable duration. If these vocal variables were achieved, talkers using a communication system under the conditions outlined by Draegert would improve their chances of being understood correctly.

There seems to be a difference of opinion in the literature on the relative merits of using the full 50-word PB lists and the 25-word half-list, that is, using either the top 25 words, or the bottom 25 words of the full 50-word PB lists. Elpern (1961) could see no reason for not employing discrimination tests which consumed less time. Hence, using all of the 24 lists in the CID W-22 series, he collected discrimination samples of 581 male patients at six Veterans Administration Audiology Clinics. Three estimates of discrimination were made:

Two were based on each of the half-lists, the third on the full 50-word list. The standard deviation was adopted as the index of stability. Since the discrepancy between half-list stability and full-list stability never exceeded 2 percent, Elpern (1961) concluded that any half-list of CID W-22 tests may be administered in lieu of the full 50-word list.

Rintelmann (1974) used CNC lists as half-lists in a clinical setting. An analysis of the results showed ". . . that half-list testing of any of these lists is warranted." Additionally, phonemic balance was questioned as an important consideration, stating ". . . equating for word familiarity rather than phonemic balance may have greater influence on list equivalency." The CNC lists also were found to have excellent test-retest reliability.

Resnick (1962) conducted a similar investigation. Using the files of the Army Audiology and Speech Center, Walter Reed Hospital, Resnick (1962) examined 51 samples each of PAL lists delivered at 30 dB above the patient's speech reception threshold. He concluded that utilization of 25-word lists would reduce the time of test administration by half, and proposed 25-word lists be used.

Campbell (1965) reported the standard deviations of Elpern's (1960) investigation were all larger than their corresponding means, making it impossible ". . . to have a value which was one standard deviation less than the mean." He also suggested the efficiency of the CID W-22 lists would be greatly improved by dropping the easier and more difficult words and replacing them with 100 words of moderate difficulty. Although he made this suggested improvement, Campbell (1965) basically felt CID W-22 word lists were ". . . inappropriate and

nonhomogeneous in word difficulty," and presented a reconstructed word list of 200 words he considered to be more homogeneous in difficulty.

In 1948 Egan stated any attempt to meet the phonetic requirements of phonetic composition, words in common usage, etc., would meet with failure. Grubb (1963) agreed. After an investigation of what happens to phonetic balance when a 50-word list is split in two, she concluded ". . . the PB characteristics of the whole list is lost in the half-list."

The relationship between the intelligibility and frequency of occurrence of English words was investigated by Howes (1957). Using a signal-to-noise ratio extending from -12 dB to +20 dB, 279 words ranging from three to 21 letters in length were presented to five college students at five different frequencies. The words selected ranged in frequency of occurrence from one to 200,000 in a sample of 4.5 million words. When correlations proved positive, Howes (1957) concluded ". . . the existence of the frequency effect for spoken words is thus confirmed." In a similar investigation, wherein distorted words were presented to normal-hearing listeners, Rosenzweig and Postman (1957) reached a similar conclusion: Word intelligibility increased directly as familiarity increased.

A study by Owens (1961) on the intelligibility of words varying in familiarity supported the investigations of Howes (1957) and Rosenzweig and Postman (1957). Owens (1961) states ". . . lists characterized by greater familiarity . . . were significantly more intelligible." According to Epstein, Giolas, and Owens (1968), the findings were related to the Harvard PB-50 and CID W-22 lists. A breakdown ". . . of the PB-50 and W-22 lists showed the markedly higher famili-

arity of the W-22 lists."

Hutton and Weaver (1959) examined variations in familiarity of the PB W-22 word lists at lower age levels. Fifteen highest rated words and 15 words rated lowest in frequency of occurrence were presented to 53 public school children, all of whom were receiving speech therapy. The words were presented at a conversational level. Results indicated the 15 least familiar words were less intelligible than the 15 most familiar words and that intelligibility increased as the age of the subjects increased. Hutton and Weaver (1959), therefore, concluded the foregoing data ". . . casts serious doubt on the use of PB W-22 words at pre-school and lower elementary levels."

Fulton (1967) used 25 normal-hearing institutionalized retardates in his investigation of the effects of practice with W-22 word lists and the word familiarity of W-22 test items. All subjects, 15 males and 10 females with a mean age of 14 years 11 months, indicated an articulation proficiency of at least 90 percent. It was found the subjects did sufficiently well with the standard W-22 lists to warrant their inclusion in test administrations. Fulton (1967), however, suggested the results also indicated List 3A to be more difficult and subjects responded better to Lists 1 and 4.

Weinhouse and Miller (1963) used 24 normal-hearing student nurses to test four versions of the Harvard PB-50 and CID W-22 word lists. Each list was presented at sensation levels of 10, 20, 30, and 40 dB. It was noted that as the sensation level increased, so did the discrimination scores. Differences between the means of the CID W-22 lists consistently yielded higher scores than the Harvard PB-50 lists. Additionally, analysis of the data indicated the difference between the

lists approached, but did not prove significant at ". . . either the 1 percent or 5 percent level of confidence at any sensation level."

Chedd (1970) writes that the greatest benefit of machines that talk will be in the computer field, where a computer will be able to speak via its speech synthesis apparatus. He also states that computers that receive Russian and translate into English already exist.

Modern computer techniques have developed three modes of interaction between man and machine: 1) computer voice readout of stored information, such as stock quotations and inventory reporting; 2) tasks involving verification of identity; and 3) automatic recognition of voiced commands. This mode can be illustrated by a conversation between an airline computer and a customer seeking flight information, then confirming (by voiced command) ticket reservations (Flanagan, 1976). The above, along with the information contained in Chapter I, would seem to indicate interest in talking machines dates back to the eighteenth Egyptian dynasty. To the best of this investigator's knowledge, however, no previous reports of empirical research of the HC 120 Phonic Mirror Handivoice are available.

CHAPTER III

METHODS AND PROCEDURES

METHODS

Subjects

Fifty-five college students ranging in age from 18 to 44 years were screened for participation in the experiment. Subjects were volunteer students attending basic communication classes at Portland State University during spring term 1980 who had been informed of the nature of the task and the approximate length of time they would be involved. They also were informed normal hearing was a prerequisite for testing the intelligibility of the instrument.

Criterion for Selection

Sole criterion for selection was normal hearing. Davis (1970) described normal hearing as the ability to hear pure tones from 500 Hz to 6,000 Hz between intensity levels of zero and 20 decibels (dB). Screening was accomplished with a Beltone Clinical Audiometer, Model 15 C, using TDH-39 earphones. Forty-eight subjects passed the hearing screening test, which was conducted in the Portland State University audiological suite.

Instrumentation

The HC 120 Phonic Mirror Handivoice (see Appendix A for specifications) is an electrically-operated speech synthesizer about the size

of a cigar box. It has a 16-button keyboard similar to those of a pocket calculator and requires punching three numbers (0-9) to produce a word or phrase. For instance, the word toothbrush is produced by punching out the following sequence: punch clear-punch enter-punch 2-punch 8-punch 2-punch talk. This required a total of six punches. Some of the spondee words used (eardrum) required as many as 25 punches. Hence, subjects could not tell whether a word or phrase was forthcoming.

Test Construction

In order to test the intelligibility of words and phrases produced by the HC 120 Phonic Mirror Handivoice, two 20-item tests were constructed. Test Number 1 was comprised of eight words each from spondee lists A, B, C, and D, and PB lists 1A, 2A, 3A, and 4A, as reproduced in Newby (1972). In addition to these 16 words, the investigator selected four phrases from the HC 120 brochure for inclusion in the test. Test Number 1 (see Appendix B), therefore, consisted of 20 scorable items.

Rationale for selection of test material was based on the fact spondee words and PB words were used not only in speech audiometry but also to assess speech processing devices (Denes and Pinson, 1973).

Test Number 2 (see Appendix C) contained eight bisyllabic and eight monosyllabic words considered to be useful in communication by this investigator. That is to say, the words were familiar and could express a mood, a need, a positive or negative reply. The test also included the same four phrases mentioned in Test Number 1. Both tests, therefore, contained 20 scorable items.

MacFarlan (1927) listed the 500 most frequently used monosyllabic words found in a compilation of four million words extracted from various literary sources. Among the 500 were: friend, no, die, which were used in this study. The list also included the words: I, need, help, how, are, me, and please. These were included in the phrases used in this study. Additionally, the word like is listed by Fletcher (1953) as one of the most frequently used words, occurring with a frequency of .11 percent. This is less than the 7.31 percent for the word the and 1.15 percent for I, but still greater than the percentages of other words.

To increase the discrimination difficulty of Test Number 2, the word bike was paired with like and included in the list of monosyllabic words. This was done to determine whether the subject could discriminate between two initial consonants, one a voiced plosive /b/, the other a voiced lingua-alveolar, lateral, non-fricative continuant /l/ (Faircloth and Faircloth, 1973).

In summary, the stimuli consisted of two tests. Test Number 1 was comprised of eight spondee words, eight PB words, and four phrases, providing 20 scorable responses. Test Number 2 was comprised of eight bisyllabic words, eight monosyllabic words, and four phrases, also providing 20 scorable responses.

Test Environment

Word intelligibility testing was done in the Portland State University audiological suite, where ambient noise level registered less than 40 dB. A General Radio Sound Level Meter, Model 1565, was used to determine suite noise level, and to calibrate the HC 120 for 60 dB

output at a distance of three feet. Maximum instrument output at that distance was approximately 72 dB. Both measurement of noise level and instrument calibration were made by a qualified audiologist.[1]

Test procedures were conducted according to the outline described by the American National Standards Institute (1971) for conducting intelligibility tests using monosyllabic words: Environmental conditions were specified, and ambient noise levels were measured with a sound level meter.

PROCEDURES

Grouping of Subjects

Eighteen males, ranging in age from 19 to 29 years with a mean age of 24 years, and 22 female students, ranging in age from 18 to 44 years with a mean age of 27.8 years, were designated as Group 1 and were given Test Number 1. Eight subjects, four male students and four female students, ranging in age from 18 to 33 years respectively, were assigned to Group 2 and were given Test Number 2. Mean age for the males was 23 years, for females 25.25 years.

Midway through the administration of Test Number 1 it became obvious scores would be low. This led the investigator to wonder whether another group, tested on a different and possibly more useful set of bisyllabic and monosyllabic words extracted from the HC 120 brochure, would do better. The same phrases used in Test Number 1 also were to be included. Hence, Test Number 2, which was to be administered

---

[1]Dr. A. Hicks, PhD, Assistant Professor of Audiology, Speech and Hearing Sciences, Portland State University.

to Group 2, was constructed.

Administration of Intelligibility Tests

Testing of instrument intelligibility was done in a sound controlled audiological suite at Portland State University. Subjects were seated three feet from the instrument and facing it. Prior to test administration they were instructed they would hear standard English words and phrases with no nonsense syllables. Some of the words would be brief in duration; hence, it was suggested they pay close attention. They were to write down what they heard.

When a long sequence of digits was required to program the HC 120, subjects had time to prepare themselves for their responses. Whenever possible, however, this examiner would punch all but the talk button of the next sequence while subjects were recording their responses. As soon as the subject stopped writing and raised his/her head from the paper, the talk button was punched, allowing subject a minimum of time between words. Tests were administered under normal speaking conditions, i.e., at a distance of three feet and at an intensity level of 60 dB (Beranek, 1954; Peterson and Gross, 1967). Overall time for Test Number 1 was approximately 10 minutes. Test Number 2, designed to require fewer punches, consumed approximately eight minutes.

Data Analysis

The overall results of Group 1 and Group 2 were reported. Mean scores, standard deviations, and range were determined for the performance of each group. Additionally, correct responses and percent correct for all categories were listed, as were syllable substitutions for

spondee and bisyllabic words, PB and monosyllabic words, and phrase substitutions.

CHAPTER IV

RESULTS AND DISCUSSION OF RESULTS

RESULTS

The present investigation sought to determine the basic intelligibility of synthesized speech produced by the HC 120 Phonic Mirror Handivoice. To accomplish this two tests, each containing 20 scorable items, were administered to two groups of normal hearing adult college students. Group 1 had 40 subjects; Group 2 had eight subjects. Test results are listed below.

Table II lists the maximum scorable responses, the mean, standard deviation, and range of both Group 1 and Group 2 and their combined scores. As can be seen from the mean and range scores of Table II, Group 2 did better overall than Group 1 in total word and phrase discrimination. Maximum scorable items, the mean, standard deviation, and range of the various stimulus types administered to Group 1 and Group 2 are listed in Table III. This table shows the test results of Group 2 bisyllabic and monosyllabic word scores were higher than those of Group 1. Additionally, because the number of phrases for both groups combined was only one-half that of other items, the mean was doubled to avoid truncated scores. Standard deviation and range were omitted for the same reason.

Correct responses and percent correct for the various stimulus types administered to Group 1 and Group 2 are posted in Tables IV

TABLE II

MEAN, STANDARD DEVIATION, AND RANGE OF OVERALL COMPOSITE SCORABLE RESPONSES FOR GROUP 1 AND GROUP 2

Source	N	Total Items	Mean	SD	Range
Group 1	40	20	7.80	2.06	3-11
Group 2	8	20	11.00	1.94	7-13
Total	48	20	8.33	2.39	3-13

TABLE III

CORRECT RESPONSES BY STIMULUS TYPE: SPONDEE WORDS, PHONETICALLY BALANCED (PB) WORDS, BISYLLABIC WORDS, MONOSYLLABIC WORDS, AND PHRASES

Type	Source	Total Items	N	Mean	SD	Range
Spondee	Test 1	8	40	2.12	1.36	0-5
PB	Test 1	8	40	2.02	0.99	0-4
Bisyl	Test 2	8	8	4.50	2.24	0-7
Monosyl	Test 2	8	8	3.12	3.59	0-8
Phrases	T1 & T2	4	48	7.24*		

*Because the number of scorable items for phrases was only one-half of all other stimulus types, the actual phrase mean was doubled in Table III to facilitate comparison. For the same reason (truncated number of scorable responses for phrases) SD and range are not reported here.

through IX. Table IV outlines the percent correct for Group 1 spondee words. This ranged from 0 percent for headlight and oatmeal to a high of 50 percent for workshop. Table V lists the correct responses and percent correct for Group 1 PB words. Ham registered 0 percent; die was correctly discriminated 72.5 percent of the time.

TABLE IV

PERCENT CORRECT FOR GROUP 1 SPONDEE WORDS
N = 40

Spondee Word	Correct Responses	% Correct
Toothbrush	14	35.0
Birthday	1	2.5
Sunset	15	37.5
Workshop	20	50.0
Headlight	0	0.0
Eardrum	16	40.0
Northwest	18	45.0
Oatmeal	0	0.0

Group 2 correct bisyllabic word responses and percent correct are indicated in Table VI. Percentage scores ranged from 0 percent for hello to 87.5 percent for sandwich. Additionally, three words from this group were correctly discriminated with a frequency of 75 percent. The number of correct responses and percent correct for Group 2 monosyllabic words are outlined in Table VII. Here it can be noted two

TABLE V

PERCENT CORRECT FOR GROUP 1 PB WORDS
N = 40

PB Word	Correct Responses	% Correct
Ran	8	20.0
Earn (urn)	14	35.0
Die (dye)	29	72.5
Ham	0	0.0
Tie	1	2.5
Chair	4	10.0
So (sew)	24	60.0
His	3	7.5

words were correctly identified with 100 percent accuracy, four were missed altogether (bread, like, bike, and fine).

Correct percentages for Group 1 and Group 2 phrases are depicted in Tables VIII and IX respectively. Percentages ranged from 72.5 percent to 97.5 percent for Group 1 phrases and from 75 percent to 100 percent for Group 2 phrases.

Syllable substitutions and frequency of substitution for Group 1 spondee words are delineated in Table X, which shows three of a possible 320 (8x40) correct first syllable interpretations, and 83 first syllable substitutions, a poor showing. Second syllable results were a little better, with 29 correct second syllable discriminations and 62 second syllable substitutions. Table XI shows there was slightly less than 50

TABLE VI

PERCENT CORRECT FOR GROUP 2 BISYLLABIC WORDS
N = 8

Word	Correct Responses	% Correct
Doctor	6	75.0
Baseball	4	50.0
Shower	2	25.0
Ice cream	5	62.5
Bathroom	6	75.0
Hungry	6	75.0
Sandwich	7	87.5
Hello	0	0.0

TABLE VII

PERCENT CORRECT FOR GROUP 2 MONOSYLLABIC WORDS
N = 8

Word	Correct Responses	% Correct
Yes	8	100.0
No	8	100.0
Bike	0	0.0
Friend	7	87.5
Rain	2	25.0
Like	0	0.0
Bread	0	0.0
Fine	0	0.0

TABLE VIII

PERCENT CORRECT FOR GROUP 1 PHRASES
N = 40

Phrase	Correct Responses	% Correct
How are you	33	82.5
I need help	29	72.5
I don't understand	37	92.5
Please leave me alone	39	97.5

TABLE IX

PERCENT CORRECT FOR GROUP 2 PHRASES
N = 8

Phrase	Correct Responses	% Correct
How are you	8	100.0
I need help	6	75.0
I don't understand	7	87.5
Please leave me alone	8	100.0

TABLE X

SYLLABLE SUBSTITUTIONS FOR GROUP 1 SPONDEE WORDS
N = 40

Word	1st Syl Correct	1st Syl Substitution	2d Syl Correct	2d Syl Substitution
Toothbrush	0	None	4	Rush (4), thrush (2), slush, ush
Birthday	0	Per (4), first (3), thurs, third	4	Se/say (10), aide (8), aid (7), stage
Sunset	0	Song (11), sound, sal	4	Says (4), said (4), that
Workshop	0	Were (4), blur, blurp, your	9	Shot, stop
Headlight	0	Said (22), ted	2	Twice (7), right (2), lice (2), ice (2), white, life's, rice
Eardrum	2	Hear (4), dear (3), tear, ther	2	Crumb
Northwest	1	Yours (6), nor (4), doors (3), nors	4	Lest (6), rest (2), left, sles, vest
Oatmeal	0	Don't (4), no (2), those, dos	0	None

percent (156/320) Group 1 PB word substitutions. This is an improvement over the spondee word results.

Group 2 bisyllabic word substitutions are outlined in Table XII, which shows only 17 word substitutions for a maximum of 64. This is a considerable improvement over spondee word results. Table XIII indicates three words had no substitutions, while one (bread) had a maximum of eight.

TABLE XI

SUBSTITUTIONS FOR GROUP 1 PB WORDS
N = 40

Word	Substitutions
Ran	Friend (13), grand (5), and (3), man (2), fran, plan, gran
Earn	Turn (6), burn (2), learn, firm, hearn
Die	High (2), I, night
Ham	Sand (7), ten (5), and (4), sam (3), tan (3), an (2), am
Chair	There (9), sayer (5), sair (3), tear (3), tare (3), share (2), szare, sare, sear
Tie	I (14), high (8), sigh (4), sighed (2), tide
So	Sold (6), sole, sawed, boat
His	This (7), is (6), says (6), six, sis, tis, fist

TABLE XII

SUBSTITUTIONS FOR GROUP 2 BISYLLABIC WORDS
N = 8

Word	Substitutions
Doctor	Lobster
Baseball	Day sol
Shower	Sour (6)
Ice cream	Light stream, stream, I strain
Birthday	None
Hello	Jello (6), cello, tello
Sandwich	None
Hungry	None

TABLE XIII

SUBSTITUTIONS FOR GROUP 2 MONOSYLLABIC WORDS
N = 8

Word	Substitutions
Yes	None
No	None
Bike	Fight (4), light, tight, sight
Friend	None
Rain	Frame, drain, thane, train
Like	Light (3), bite, tight, right, night
Fine	Sign (4), sigh (3), find
Bread	Thread (8)

Rearrangements, rather than substitutions of combined Group 1 and Group 2 phrases are shown in Table XIV. I need help had the most; please leave me alone had the least.

TABLE XIV

SUBSTITUTIONS FOR GROUP 1 AND GROUP 2 PHRASES
N = 48

Phrase	Phrase Substitutions
I need help	He needs help (2), all he needs is help, Holly needs help, guy needs help, Sally needs help, ale needs help, I need shelves
How are you	So how are you (2), Sir, How are you
I don't understand	Have you understand, Sally don't understand
Please leave me alone	Leave me alone

DISCUSSION OF RESULTS

On examining the mean and ranges of Table II, it becomes discernible Group 2 did better overall than Group 1. This again becomes evident when one examines the Group 2 bisyllabic and monosyllabic word means in Table III. The mean for combined Group 1 and Group 2 phrases, also contained in Table III, was much higher than the mean for correct responses on all categories of single word responses. This was probably due to the fact phrases have more contextual cues than do isolated words. Additionally, the majority of the words in the phrases was considered by MacFarlan (1927) to be among the 500 most frequently used

monosyllabic words in the English language.

Table IV indicates the subjects of Group 1 failed to discriminate the spondee words headlight and oatmeal, both words registering 0 percent correct. The initial consonant /h/ of headlight is a glottal fricative-continuant, and although voiceless, requires pressure to enunciate. Additionally, the phoneme /h/ functions ". . . only as a syllable releasing consonant (Faircloth and Faircloth, 1973)." As will be seen in the forthcoming discussion, this may have been the reason that, with only one exception, all words in this study beginning with the initial consonant /h/ proved difficult to discriminate. The word oatmeal as produced by the HC 120, was a little "ragged" and seemed to be emitted in three syllables instead of two. Birthday was correctly identified only once in 40 Group 1 test administrations.

The remaining spondee words were fairly well grouped. Workshop, however, scored highest, with a correct score of 50 percent. The most frequent first syllable substitution in birthday (4) was the voiceless, bilabial, plosive /p/ for the voiced, bilabial, plosive /b/. The next most frequent first syllable substitution (3) was the labiodental, fricative-continuant /f/ (see Table X). The onus for this poor showing then, may have been in the HC 120's production of birthday. Workshop contained two consonants, /r/ and /k/, which were ranked sixth and ninth (of 25) in the order of frequency of occurrence of consonants in the connected speech of normal speakers by Faircloth and Faircloth (1973). Frequency of occurrence leads to familiarity, which in turn leads to better understanding of the spoken word. Hence, the phoneme combination /rk/ probably enabled subjects to discriminate workshop with a fairly high degree of accuracy.

A glance at Table V, which lists the correct percentages for PB words, reveals one word (ham) was missed altogether, one word (tie) was correctly discriminated once in 40 test administrations, and the word his elicited 7.5 percent correct responses. Table XI indicates the initial /t/, a voiceless, lingua-alveolar, plosive (Faircloth and Faircloth, 1973), was dropped from the word tie 14 of 40 test administrations, indicating either poor production by the HC 120 or poor discrimination on the part of the subjects. The words ham and his, like headlight, began with the consonant /h/, which seemed difficult for the HC 120 to produce.

Die was correctly interpreted 72.5 percent of the time. This is in keeping with the literature (MacFarlan, 1927), which lists die as one of the 500 most frequently used monosyllabic words. Good production by the word die by the HC 120 enabled subjects to achieve a high discrimination score.

Only one word, hello, had a 0 percentage score as noted in Table VI, which lists the discrimination percentages for Group 2 bisyllabic words. Table XII shows jello as the most frequent substitution for hello, the fourth word wherein the initial consonant /h/ was poorly discriminated. The word hungry, however, was more familiar and could be understood without the initial /h/. If a small child were to say "m'ungry," the expression would more than likely be translated as "I'm hungry." This may be due to the fact the word is a very familiar one, or that there is a slight pause between the two syllables. The results are in keeping with the findings of Owens (1961), who states: ". . . lists characterized by greater familiarity . . . were significantly more intelligible." Hence, the word hungry was interpreted correctly

with a frequency of 75 percent. Subjects reported hungry correctly or not at all. Additionally, Table VI shows the words doctor and bathroom also were reported with 75 percent accuracy, while sandwich scored a high of 87.5 percent.

It must be remembered Group 2 bisyllabic and monosyllabic words were chosen by this investigator on the basis of their familiarity and their usefulness to the handicapped individual. That is to say, the words could express a need, a mood, a positive or negative reply in one word. This may account for the high degree of accuracy of Group 2 monosyllabic words contained in Table VII.

Table VII lists the correct percentages for Group 2 monosyllabic words, which were evenly split between words correctly discriminated and those not discriminated. Yes and no, being more familiar and having greater usage than other monosyllabic words, were discriminated at the 100 percent level. Close behind was friend with an 87.5 percent frequency score, which was more or less expected since the word was listed by MacFarlan (1927) as being one of the 500 most frequently used monosyllabic words. High frequency or not, this in no way detracts from the reputable task accomplished by the HC 120 in the production of the three words just under discussion.

Rain was correctly identified with a frequency of 25 percent. Bike, like, bread, and fine registered 0 in the percent column. Bike was paired with like for two reasons: Fletcher (1953) considered like to be one of the most frequently used words (see Chapter III). Would bike be as accurately interpreted? If not, would it be due to the initial consonants of the two words? In other words, would subjects be able to discriminate between /l/ and /b/? We see from Table XIII

the most frequent substitution for bike was fight, wherein a voiceless, labiodental, fricative-continuant /f/ was substituted for the voiced, bilabial, plosive /b/. The initial consonant /l/ of like was correctly discriminated three times out of eight. The phoneme /k/, however, was heard as a /ght/ sound, indicating the substitution of a lingua-alveolar, voiceless, plosive /t/ for a lingua-alveolar, plosive /k/ (Faircloth and Faircloth, 1973).

It is difficult to explain why like was so poorly discriminated, particularly in view of its familiarity. One can only conclude the HC 120 did not produce a reasonable facsimile of the three phonetic elements in the word like. It seems reasonable to assume bike was probably poorly discriminated for the same reason.

According to the contents of Table XIII, the word bread was consistently recorded as thread. Since the error was 100 percent, it can only be assumed the instrument did not produce a clear initial consonant. Although the vowel sound /aɪ/ of fine was correctly discriminated, the initial consonant /f/ was interpreted as the lingua-alveolar, voiceless, fricative-continuant /s/ (Faircloth and Faircloth, 1973). Again, the high percentage of error leads one to conclude the HC 120 was not clear in the production of an initial /f/.

Correct percentages for Group 1 and Group 2 phrases are outlined in Tables VIII and IX. These tables indicate that Group 2 did somewhat better overall than Group 1. The phrase scores for both groups were, in general, much higher than those of other stimulus types. This is not surprising, since phrases offer more contextual cues than do single isolated words. Additionally, the words used in the phrases were considered by MacFarlan (1927) to be among the 500 most frequently used

monosyllabic words. With respect to this latter point, however, like also was a high frequency word (Fletcher, 1953), occurring with a frequency of .11 percent among the single words presented, yet it achieved 0 percent correct.

Syllable substitutions for Group 1 spondee words are delineated in Table X. Regarding first syllable performance, there were two words, eardrum and northwest that elicited correct first syllable responses. The number correct, however, was negligible. The same holds true for the second syllable. With the exception of workshop, no item had more than 10 percent. Toothbrush, which had 14 correct responses, or 35 percent correct, produced eight second syllable substitutions, all of which were homonyms for brush. Birthday, 2.5 percent correct, had 10 first syllable substitutions of various sorts, and 26 second syllable substitutions, 10 of which contained the same diphthong /eɪ/ as day. There were 11 substitutions of song for sun in sunset and nine second syllable substitutions, eight of which began with /s/. Workshop had four different first syllable substitutions for a total of seven, and two second syllable substitutions.

The substitutions above have led this investigator to the hypothesis the HC 120 was not designed as a testing vehicle for spondee and PB words, and should not be considered as such. Its main function was to provide handicapped individuals with a means of communication. This hypothesis seems to be reinforced by the excellent results obtained in phrase responses, all of which were organic to the HC 120. Bearing this in mind, we continue to discuss the remaining spondee words, and the PB list of substitutions.

The most frequent first syllable substitutions occurred in the word headlight, which had 22 substitutions of said for head and five different initial second syllable substitutions for a total of 16. Eardrum had nine first syllable substitutions divided among four different initial syllables, and one second syllable substitution. Considering the fact eardrum required 25 "punches" on the HC 120, this is not too surprising. Northwest had 14 first syllable substitutions distributed among four words, and 11 second syllable substitutions beginning with four different consonants, while oatmeal had a variety of first syllable substitutions totalling eight, and no second syllable substitutions.

Substitutions for Group 1 PB words are outlined in Table XI. Of the 23 substitutions recorded for ran, 13 were an unlikely friend, indicating poor production by the HC 120. Earn had 11 substitutions, all of which could be considered homonyms. Here an initial consonant was inserted before the stressed /r/ of earn. Die, ham, chair, tie, so, and his had 4, 25, 28, 29, 9, and 23 substitutions respectively. Group 2 bisyllabic word substitutions, as listed in Table XII, were few, with hello and shower each registering six. Shower was recorded as sour indicating a substitution of /s/ for /ʃ/ in the initial syllable. This would tend to indicate the HC 120's production of /ʃ/ was incomplete.

The combined results of Group 1 and Group 2 phrase substitutions in Table XIV indicate two of the three words of I need help were correctly identified. How are you had two additions of so and one of sir preceding the phrase. The question arises, did subjects really hear

so and sir. If so, why were not more recorded? I don't understand was recorded as have you understand and Sally don't understand. These misinterpretations are difficult to explain. The word please was omitted from the phrase Please leave me alone, more than likely through inattention.

An overview of the results indicates that overall, Group 2 had the highest mean and range scores, and the highest bisyllabic word mean. Correct responses to the various stimuli, in descending order of performance, were as follows: Group 2 phrases, Group 1 phrases, Group 2 bisyllabic words and monosyllabic words, Group 1 spondee words, and Group 1 PB words.

Success of phrases, which elicited a notably higher percentage of correct responses, ranged from a low of 72.5 percent to a high of 97.5 percent for Group 1 and from a low of 75 percent to a high of 100 percent for Group 2.

This was not surprising since the phrases were organic to the HC 120, and the words contained therein were considered to be among the 500 most frequently used monosyllabic words in the English language (MacFarlan, 1927), and contained more contextual cues than single isolated words. Opposed to this was the difference of opinion in the literature on the purpose for construction of spondee and PB word lists. Hirsh et al. (1952) emphasized familiarity in the spondee word lists and the degree of difficulty among words in the W-22 word lists. Lehiste and Peterson (1959) considered phonetic balance important and developed a test consisting of CNC type words.

There also was disagreement among investigators as to whether a full 50-word list was really necessary. Elpern (1961) favored employ-

ment of a time-saving half-list of 25 words. Campanelli (1962), Resnick (1962), and Rintelmann (1974) agreed. They were opposed by Egan (1948), who claimed use of 25 words would not meet the requirements of phonetic composition, usage, etc. Grubb (1963) agreed. She investigated phonetic balance when a full 50-word list was split in two and concluded: ". . . the PB characteristics of the whole list is lost in the half-list.

The total set of words and phrases was carefully selected (as discussed in Chapter III) to assess the performance of the HC 120 across 48 subjects. Both the number of test items and the wide variability of item content were deemed sufficient for this purpose.

CHAPTER V

SUMMARY AND IMPLICATIONS

SUMMARY

This study sought to determine the basic intelligibility of synthesized speech as produced by the HC 120 Phonic Mirror Handivoice. It involved 48 male and female subjects divided into two groups. Group 1 had 40 subjects ranging in age from 18 to 44 years. Group 2 had eight subjects ranging in age from 18 to 33 years. All subjects were screened for normal hearing. Testing for normal hearing was done in an audiological suite at Portland State University. The study was designed to determine the degree to which a group of normal hearing college students could recognize words and phrases as produced by the HC 120 Phonic Mirror Handivoice.

In order to test the intelligibility of words and phrases produced by the HC 120 Phonic Mirror Handivoice, two 20-item tests were constructed. Test Number 1 was comprised of eight spondee words, eight phonetically balanced (PB) monosyllabic words, and four phrases taken from the HC 120 brochure for a total of 20 scorable items. Test Number 2 contained eight bisyllabic words, eight monosyllabic words contained in the HC 120 brochure, and the same four phrases mentioned in Test Number 1. Both tests, therefore, contained 20 scorable items.

Results indicated that with the exception of hungry, all words (ham, headlight, his, hello) beginning with the consonant /h/ were

difficult to discriminate. Two spondee words, oatmeal and birthday, were very poorly discriminated, as were the PB words tie and chair. Monosyllabic words were evenly divided between those discriminated and those not discriminated. The initial consonants of like and bike were not discriminated and neither were the words. Bread was consistently identified as thread, while fine was not discriminated at all.

Phrase identification for both Group 1 and Group 2, containing more contextual cues, was very well discriminated. It would seem then, the HC 120 had limited capabilities in the production of isolated words. As far as words and phrases organic to itself, the HC 120 had good to excellent capabilities.

IMPLICATIONS

To the best of this writer's knowledge, no previous study has been done on the intelligibility of the HC 120 Phonic Mirror Handivoice; hence, comparisons for use with the present study were not available. It is recommended, however, intelligibility tests using the HC 120 should be replicated. Spondee words, however, should be eliminated since their production sometimes requires many "punches" to produce.

Two test conditions are suggested: Condition A: Present words and phrases to a heterogeneous group of subjects, minus spondee words, with the HC 120. Condition B: Present the same stimuli at a later date, using a tape-recorded version of the word and phrase list. Compare the results. Having a practiced and non-practiced group could add another dimension to the test.

One also could construct a full 50-word bisyllabic or monosyl-

labic word list and a 25-word half-list, and compare results when one set is produced by the HC 120, the other by a tape-recorded reproduction of a voiced reading by experienced talkers. This would provide a basis for comparison with studies using full 50-word lists and 25-word half-lists.

Since rate is considered important by some researchers, words and/or words and phrases produced by both the HC 120 and experienced talkers could be timed with a Kay Sona-graph. Comparisons, not only of the time involved, but also of the different sonagraphic reproductions of the words and/or words and phrases could be made. It is felt this last suggestion would point out differences, if any, in word/phrase production, and provide HC Electronics with an indication where improvement would be useful.

We note several references have been made to contextual cues. These are important to speech perception. Perkins (1971) states nearly all sounds of a syllable are influenced by context, as well as by other parameters. Hence, contextual cues are important to understanding utterances, whether they be isolated words or phrases. Thus, this investigator suggests that a study comparing contextual cues may be of value. In fact, contextual cues may explain why phrases consistently scored higher than isolated words.

Finally, it is the opinion of this investigator that in time, improvements in producing synthesized speech will engender increased intelligibility. The HC 120 Phonic Mirror Handivoice is a talking machine with great potential.

SELECTED BIBLIOGRAPHY

AMERICAN NATIONAL STANDARDS INSTITUTE, Method for measurement of monosyllabic word intelligibility. American National Standards Institute S3.2 (1971).

ANONYMOUS, Pedro the Voder. Bell Telephone Record 17, 170-183 (1939).

BENSON, R. W., Davis, H., Harrison, C. E., Hirsh, I. J., Reynolds, E. G., and Silverman, S. R., C.I.D. Auditory Tests W-1 and W-2. J. Acoust. Soc. Amer. 23, 719 (1951).

BERANEK, L. L., Acoustics. New York: McGraw-Hill (1954).

BLACK, J. W., Multiple choice intelligibility tests. J. Speech and Hearing Dis. 22, 213-235 (1957).

BUNCH, C. C., The development of the audiometer. Laryngoscope 57, 57-89 (1947).

CAMPANELLI, P. A., A measure of intra-list stability of four PAL word lists. J. Auditory Res. 2, 50-55 (1962).

CAMPBELL, R. A., Discrimination word test difficulty. J. Speech and Hearing Res. 8, 13-22 (1965).

CHEDD, G., Sound. Garden City, N.Y.: Doubleday and Co. (1970).

DAVIS, H., Acoustics and psychoacoustics. In H. Davis and S. R. Silverman: Hearing and Deafness. New York: Holt, Rinehart, and Winston (1970).

DENES, P. B., and Pinson, E. N., The Speech Chain: The Physics and Biology of Spoken Language. Garden City, N.Y.: Doubleday and Co. (1973).

DRAEGERT, G. I., Relationships between voice variables and speech intelligibility in high level noise. Speech Monographs 18, 272-278 (1951).

DUDLEY, H., and Tarnoczy, T. H., The speaking machine of Wolfgang von Kempelen. J. Acoust. Soc. Amer. 22, 151-167 (1950).

EGAN, J. P., Articulatory testing methods. Laryngoscope 9, 955-991 (1948).

ELPERN, B. S., Differences in difficulty among the CID W-22 auditory tests. Laryngoscope 70, 1560-1566 (1960).

ELPERN, B. S., The relative stability of half-list and full list discrimination tests. Laryngoscope 71, 30-36 (1961).

ENCYCLOPEDIA OF SCIENCE AND TECHNOLOGY, vol. 14. New York: McGraw-Hill (1977).

EPSTEIN, A. Giolas, T. G., and Owens, E., Familiarity and intelligibility of monosyllabic word lists. J. Speech and Hearing Res. 11, 435-438 (1968).

FAIRCLOTH, S. R., and Faircloth, M. A., Phonetic Science: A Program of Instruction. Englewood Cliffs, N.J.: Prentice-Hall (1973).

FLANAGAN, J. L., Computers that talk and listen: Man-machine communication by voice. Proceedings of the IEEE 64, 405-415 (April 1976).

FLETCHER, H., Speech and Hearing. New York: Van Nostrand Co. (1950).

FLETCHER, H., Speech and Hearing in Communication. Princeton, N.J.: Van Nostrand Co. (1953).

FULTON, R. T., Task adaptation and word familiarity of W-22 word discrimination lists with retarded children. J. Auditory Res. 7, 353-358 (October 1967).

GIOLAS, T. G., and Epstein, A., Comparative intelligibility of word lists and continuous discourse. J. Speech and Hearing Res. 6, 349-358 (1963).

GLORIG, A., Audiometry: Principles and Practices. Baltimore, Md.: Williams and Wilkins Co. (1965).

GOETZINGER, C. P., Word discrimination testing. In J. Katz (Ed.): Handbook of Clinical Audiology. Baltimore, Md.: Williams and Wilkins Co. (1972).

GRUBB, P., A phonemic analysis of half-list speech discrimination tests. J. Speech and Hearing Res. 6, 271-273 (1963).

HC ELECTRONICS CORPORATION, HC 120 Phonic Mirror Handivoice brochure. Mill Valley, Calif.: (1979).

HIRSH, I. J., Silverman, S. R., Reynolds, E. G., Eldert, E., and Benson, R. W., Development of materials for speech and audiometry. J. Speech and Hearing Dis. 17, 321-337 (September 1952).

HOWES, D., On the relation between the intelligibility and frequency of occurrences of English words. J. Acoust. Soc. Amer. 29, 296-305 (1957).

HUDGINS, C. V., Hawkins, J. E., Karlin, J. E., and Stevens, S. S., The development of recorded auditory tests for measuring hearing loss for speech. Laryngoscope 57, 57-89 (1947).

HUTTON, C., and Weaver, J., PB intelligibility and word familiarity. Laryngoscope 69, 1443-1450 (1959).

LEHISTE, I., and Peterson, G. E., Linguistic considerations in the study of speech intelligibility. J. Acoust. Soc. Amer. 31, 280-286 (1959).

MacFARLAN, D., The voice test of hearing. Archives Otolaryn. 5, 1-30 (1927).

MOSES, E. R., Phonetics: History and Interpretation. Englewood Cliffs, N.J.: Prentice-Hall (1964).

NEWBY, H. A., Audiology. New York: Appleton-Century-Crofts, Inc. (1972).

OWENS, E., Intelligibility of words varying in familiarity. J. Speech and Hearing Dis. 4, 113-129 (1961).

PERKINS, W. H., Speech Pathology: A Behavioral Science. St. Louis, Mo.: C. V. Mosby Co. (1971).

PETERSON, P. G., and Gross, E. E., Handbook of Noise Measurement. West Concord, Mass.: General Radio Co. (1967).

PICKETT, J. M., and Pollack, I., Intelligibility of excerpts from fluent speech: Effects of rate of utterance and duration of excerpt. Language and Speech 6, 151-164 (1963).

POLLACK, I., and Pickett, J. M., The intelligibility of excerpts from conversation. Language and Speech 6, 165-171 (1963).

READ, O., and Welch, W. L., From Tinfoil to Stereo. New York: Bobbs-Merrill Co. (1959).

RESNICK, D. M., Reliability of the 25-word phonetically balanced lists. J. Auditory Res. 2, 5-12 (1962).

RINTELMANN, W. F., Six experiments on speech discrimination. J. Auditory Res. Supp. 2 (1974).

ROBINSON, A. L., More people are talking to computers as speech recognition enters the real world. Science 203, 634-638 (1979).

ROSENZWEIG, M. R., and Postman, L., Intelligibility and frequency of usage. J. Exper. Psych. 54, 412-422 (1957).

SCHROEDER, M. R., Vocoders: Analysis and synthesis of speech. Proceedings of the IEEE 54, 720-734 (1966).

WEINHOUSE, I., and Miller, M. H., Discrimination scores for two lists of phonetically balanced words. J. Auditory Res. 3, 9-14 (1963).

APPENDIX A

HC 120 PHONIC MIRROR HANDIVOICE SPECIFICATIONS

The HC 120 Phonic Mirror Handivoice is a communication prosthesis, a diagnostic instrument, an educational and therapy tool that looks and operates like a calculator. All selections are accomplished through a three-digit numeric coding. It is programmed with 893 words, 45 phonemes to create any word, all 26 letters of the alphabet, 13 morphemes, and 16 short phrases.

The instrument features a 16-button keyboard, has six selectable functions, including memory and repeat modes, audible and silent automatic storage scanning when used with auxiliary controls. It has auxiliary breath, muscle or hand switch controls to accommodate the severely physically impaired, maximizing their available motor skills. The HC 120 also provides unlimited vocabulary through the combined use of morphemes, phonemes, letters, and words.

Information on electronic components were not available.

APPENDIX B

TEST 1 WORDS AND PHRASES

Spondee Words	PB Words	Phrases
Toothbrush	Ran	How are you
Birthday	Earn (urn)	I need help
Sunset	Die (dye)	I don't understand
Workshop	Ham	Please leave me alone
Headlight	Tie	
Eardrum	Chair	
Northwest	So (sew)	
Oatmeal	His	

APPENDIX C

TEST 2 WORDS AND PHRASES

Bisyllabic Words	Monosyllabic Words	Phrases
Doctor	Yes	How are you
Baseball	No (know)	I need your help
Shower	Bike	I don't understand
Ice cream	Friend	Please leave me alone
Bathroom	Rain	
Hungry	Like	
Hello	Bread	
Sandwich	Fine	