Intelligibility of Speech Produced with an Artificial Larynx at Various Frequencies

Sandra Arlene Merritt

Western Michigan University

Follow this and additional works at: https://scholarworks.wmich.edu/masters_theses

Part of the Communication Sciences and Disorders Commons

Recommended Citation
Merritt, Sandra Arlene, "Intelligibility of Speech Produced with an Artificial Larynx at Various Frequencies" (1982). Master's Theses. 1672.
https://scholarworks.wmich.edu/masters_theses/1672
INTELLIGIBILITY OF SPEECH PRODUCED WITH
AN ARTIFICIAL LARYNX AT VARIOUS FREQUENCIES

by

Sandra Arlene Merritt

A Thesis
Submitted to the
Faculty of The Graduate College
in partial fulfillment of the
requirements for the
Degree of Master of Arts
Department of Speech Pathology and Audiology

Western Michigan University
Kalamazoo, Michigan
August 1982
INTELLIGIBILITY OF SPEECH PRODUCED WITH
AN ARTIFICIAL LARYNX AT VARIOUS FREQUENCIES

Sandra Arlene Merritt, M.A.

Western Michigan University, 1982

Word intelligibility scores of 21 listeners were used to test the hypothesis that speech intelligibility will vary systematically across speakers and fundamental frequency of electrolarynx vibration. Twenty-one listeners transcribed audio recorded lists of CVC utterances produced with an electrolarynx by 3 speakers at each of five fundamental frequencies.

Comparative analysis of scores across frequencies and speakers were discussed in terms of:

a) source transmission characteristics of "electrolaryngeal" speech, b) implications for diagnosis/therapeusis of alaryngeal speakers.

Results indicated that intelligibility scores were significantly different between speakers and between frequencies within speakers. Maximum intelligibility scores were more highly correlated with certain frequencies for certain speakers. Results indicated that the relation of maximum intelligibility scores and fundamental frequency of vibration is speaker dependent.

Additional post hoc information about phoneme error type substitutions, sex recognition, and quality judgments of acceptability of the speakers was related to intelligibility differences.
ACKNOWLEDGEMENT

I would like to thank the three surviving members of my thesis committee: John Hanley, Robert Erickson, and Bill Dawson. A special thanks to my thesis advisor, John "Mick" Hanley, for his time, effort and unending patience. His instruction and guidance over the past year has made this learning experience a valuable one. Thanks to Robert Erickson for his generosity of clinical equipment and personal references. Thanks also to Bill Dawson for his technical assistance, continuous encouragement, and for demonstrating a genuine interest in this study.

Appreciation is extended to the subjects who volunteered for the study, especially Carol Heermann-Jones, Carol French, and Gary Westra.

Finally, a personal thanks to Craig Buffie, whose presence in my life has made many things possible - and doing them seem worthwhile.

Sandra Arlene Merritt
INFORMATION TO USERS

This reproduction was made from a copy of a document sent to us for microfilming. While the most advanced technology has been used to photograph and reproduce this document, the quality of the reproduction is heavily dependent upon the quality of the material submitted.

The following explanation of techniques is provided to help clarify markings or notations which may appear on this reproduction.

1. The sign or "target" for pages apparently lacking from the document photographed is "Missing Page(s)". If it was possible to obtain the missing page(s) or section, they are spliced into the film along with adjacent pages. This may have necessitated cutting through an image and duplicating adjacent pages to assure complete continuity.

2. When an image on the film is obliterated with a round black mark, it is an indication of either blurred copy because of movement during exposure, duplicate copy, or copyrighted materials that should not have been filmed. For blurred pages, a good image of the page can be found in the adjacent frame. If copyrighted materials were deleted, a target note will appear listing the pages in the adjacent frame.

3. When a map, drawing or chart, etc., is part of the material being photographed, a definite method of "sectioning" the material has been followed. It is customary to begin filming at the upper left hand corner of a large sheet and to continue from left to right in equal sections with small overlaps. If necessary, sectioning is continued again—beginning below the first row and continuing on until complete.

4. For illustrations that cannot be satisfactorily reproduced by xerographic means, photographic prints can be purchased at additional cost and inserted into your xerographic copy. These prints are available upon request from the Dissertations Customer Services Department.

5. Some pages in any document may have indistinct print. In all cases the best available copy has been filmed.

University Microfilms International
300 N. Zeeb Road
Ann Arbor, MI 48106

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
MERRITT, SANDRA ARLENE

INTELLIGIBILITY OF SPEECH PRODUCED WITH AN ARTIFICIAL LARYNX AT VARIOUS FREQUENCIES

WESTERN MICHIGAN UNIVERSITY M.A. 1982

Copyright 1982
by

MERRITT, SANDRA ARLENE

All Rights Reserved

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
# TABLE OF CONTENTS

ACKNOWLEDGMENTS ........................................ ii

LIST OF TABLES ........................................ vi

CHAPTER

I. BACKGROUND AND PURPOSE ......................................... 1
   Introduction ........................................ 1
   Statement of the Problem ................................ 17

II. METHOD AND PROCEDURES ........................................ 19
   Speakers ........................................ 19
   Stimulus Materials ..................................... 20
   Apparatus ........................................ 20
   Recording and Presentation of Stimulus Words .......... 22

III. RESULTS ...................................................... 25
   The Relationship Between Fundamental Frequencies
   (F₀) of Electrolarynx Vibration and Speech
   Intelligibility Scores for Speakers ..................... 28
   Analysis of Phoneme Error by Frequency ............. 37
   Summary ........................................ 44

IV. DISCUSSION .................................................... 46
   The Relationship Between Fundamental Frequency
   (F₀) of Electrolarynx Vibration and Speech
   Intelligibility Scores for Speakers ..................... 47
   Analysis of Phoneme Errors by Frequency ............. 54

V. SUMMARY AND CONCLUSIONS ...................................... 67
   Purpose ........................................ 67
   Experimental Design ..................................... 67
   Findings and Conclusions ................................ 69

iii

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
APPENDICES. .................................................. 72

A. PB-50 Word Lists Selected From the USA Monosyllabic
   Intelligibility Test Word Lists. .................. 73

B. Schematic of Adjusted Electrolarynx
   Modified Switch Circuitry. ..................... 76

C. Sound Pressure Levels for Unloaded
   and Loaded Systems at Five Test
   Frequencies. ........................................... 77

D. Recording Order for Word Lists (1, 2, 3, 4, 5)
   and Frequencies 94 Hz. (A), 120 Hz. (B), 160 Hz.,
   (C), 212 Hz. (D), 245 Hz. (E) for Speaker 1,
   Speaker 2, and Speaker 3 ......................... 78

E. Rank Order of Intelligibility Scores and
   Acceptability Ratings (Low to High) and Order
   of Frequency Presentation for Each Group and Speaker.
   Frequency A = 94 Hz., Frequency B = 120 Hz.,
   Frequency C = 160 Hz., Frequency D = 212 Hz.,
   and Frequency E = 245 Hz. ....................... 79

G. Sound Spectra of the Vowel Sound "a" (as in
   "father") Produced With the Electrolarynx
   Set at the Five Test Frequencies ............... 81

F. Sound Spectra of the Vowel "ee" (as in "bean")
   Produced With the Electrolarynx Set at the
   Five Test Frequencies. .............................. 80

H. Sound Spectra of the Electrolarynx Set at the
   Five Test Frequencies, 94 Hz., 120 Hz., 160 Hz.,
   212 Hz., 245 Hz. ................................. 82

I. Rank Order of Mean Intelligibility Scores
   (Parenthesis) From Low to High and Their Corresponding
   Frequencies For Each Speaker Across Groups
   Frequency A = 94 Hz., B = 120 Hz., C = 160 Hz.,
   D = 212 Hz., and E = 245 Hz. .................... 83

J. Instructions For Speakers and Listeners. ...... 84

iv

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
# TABLE OF CONTENTS — CONTINUED

## APPENDICES

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>K. Informed Consent Release Form.</td>
<td>86</td>
</tr>
<tr>
<td>L. Listeners Response Sheet</td>
<td>87</td>
</tr>
<tr>
<td>M. Matrix of Percentage of Total Phoneme Error Substitutions by Speaker: Percentages of Frequencies A(94Hz.), B(120Hz.), C(160Hz.), D(212Hz.), E(245Hz.) are Listed Sequentially</td>
<td>88</td>
</tr>
</tbody>
</table>

## BIBLIOGRAPHY

91
LIST OF TABLES

1. Intelligibility Scores (Percentages) for Each Speaker At Each of Five Test Frequencies (A=94Hz., B=120Hz., C=160Hz., D=212Hz., and E=245Hz.) for Each Listener For Three Groups of Listeners ........................................... 26

2. Individual and Pooled Group Mean Intelligibility Scores for Each Speaker Across Frequencies ........ 27

3. Pooled Mean Intelligibility Scores Across Frequencies for Speaker 1(S1), Speaker 2 (S2) and Speaker 3(S3). 29

4. Rank Order of Intelligibility (Low-High) of Test Frequencies For Listening Groups. Frequency A=94Hz., Frequency B=120Hz., Frequency C=160Hz., Frequency D=212Hz., Frequency E=245Hz. Lines Below Frequency Orders Depict Intelligibility Scores for Frequencies Which Were Not Significantly Different (p<.05) ................. 30

5. Mean Intelligibility Scores and Standard Deviations for Speaker 1 Across Frequencies and Groups. ...... 33

6. Mean Intelligibility Scores and Standard Deviations for Speaker 2 Across Frequencies and Groups. ........ 34

7. Mean Intelligibility Scores and Standard Deviations for Speaker 3 Across Frequencies and Groups. ........ 35

8. Pooled Group Mean Intelligibility Scores For Speakers at Each Frequency ............................... 38

9. Matrix of Total (577) Phoneme Substitution Errors and Percentage of Total Phoneme Errors (Parenthesized) for Speaker 1 Across All Frequencies and Listeners. Only Substitution Types Accounting for More than 1% of the Error Total are Presented. ................................................. 40

10. Matrix of Total (851) Phoneme Substitution Errors and Percentage of Total Phoneme Errors (Parenthesized) For Speaker 2 Across All Frequencies and Listeners. Only Substitution Types Accounting for More Than 1% of the Error Total are Presented ............ 41

11. Matrix of Total (1109) Phoneme Substitution Errors and Percentages of Total Phoneme Errors (Parenthesized) for Speaker 1 Across All Frequencies and Listeners. Only Substitution Types Accounting for More Than 1% of the Error Total are Presented ............ 42
LIST OF TABLES — CONTINUED

12. Percentage of Vowel Substitution Errors For Speakers Occurring at Each Frequency ........ 45

13. Rank Order of Intelligibility Scores and Acceptability Ratings (Low-High) and Order of Frequency Presentation for Each Group and Speaker Frequency A=94Hz., Frequency B=120Hz., Frequency C=160Hz., Frequency D=212Hz., Frequency E=245Hz. ........ 49

14. Percentages of Listeners in Each Group That Indicated the Speaker was Male. ............. 63
CHAPTER 1

BACKGROUND AND PURPOSE

Introduction

A number of compensatory communication modes are used by individuals to cope with various speech disorders. The laryngectomized person usually implements one of two alternative modes, esophageal speech or an artificial larynx. While the use of a prosthetic voicing source or artificial larynx is the second most common form of alaryngeal speech, its use is becoming more commonplace. Weinberg (1980) reports that there are two principal reasons for the widespread and increasing use of artificial larynges; 1) patient preference for the artificial larynx as a primary method of oral communication and 2) failure of large numbers of laryngectomized patients to attain functional esophageal speech. It is estimated that one-third of the laryngectomized population are physically incapable of utilizing esophageal speech as a primary means of communication (Lauder, 1968). Historically, the artificial larynx was commonly regarded as a back-up method for patients who failed to achieve satisfactory esophageal speech. However, researchers have recently begun arguing that a primary goal of alaryngeal speech rehabilitation was to develop functional communication, regardless of the source mode of voicing (Diedrich & Youngstrom, 1966). The Diedrich concept, coupled with the fact that there is no evidence indicating that the use of an artificial larynx reduces the laryngectomized patient's ability to learn esophageal speech.

1 Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
speech, gave rise to increasing interest in the use of artificial larynges as an alternative and primary mode of voicing support for speech (Weinberg, 1980). This interest led to a currently popular clinical philosophy which espouses the use of the artificial larynx as a compliment to esophageal voice training. Similarly, use of the artificial larynx may compliment the use of other prosthetic devices or surgical procedures which have recently been developed (Singer and Blom, 1980). According to Lauder (1968), this strategy is sound for a variety of reasons. The early introduction and use of the artificial larynx following laryngectomy is a psychological as well as an economic necessity for the new laryngectomee. Esophageal voice can often be developed more readily when the patient has recovered from this traumatic surgical experience. The immediate post-trauma period may be best suited to the use of the electrolarynx. Specialists also claim that the artificial larynx is much more understandable than esophageal voice. It enables the user to communicate more effectively soon after surgery, particularly in situations involving emotional stress or in situations which require a speech signal which is more intense than is normally produced with esophageal voice (Lauder, 1968). Reports and clinical observations also indicate that persons articulate well esophageally if they have successfully used the artificial larynx before acquiring esophageal speech (Lauder, 1968). Others, (Diedrich and Youngstrom, 1966) suggested that the artificial aid actually hastens the learning of esophageal speech by facilitating improvement of articulatory skills. The user of the artificial larynx must articulate precisely to achieve intelligible speech.
He must learn, for example, to make voiceless consonant sounds by developing adequate intraoral and pharyngeal air pressures to compensate for the loss of pulmonary air. The learner of esophageal phonation must also learn to articulate voiceless consonants in a similar compensatory manner.

Unlike esophageal voicing, the use of an artificial larynx is not dependent on the integrity of the esophagus and the pharyngeal-esophageal segment. The vibrating source of these artificial sound devices are powered either by pulmonary air or by electronic means. An artificial larynx driven by pulmonary air utilizes air expelled from the stoma to activate the vibrating source. These devices are not commonly used. The electronic device generates a sound into the vocal tract by either placing the vibrating diaphragm against the neck or by intra-oral insertion. The vibrating source, when coupled with the pharyngeal, nasal, and oral cavities, produces acoustic spectra similar to speech produced by the human larynx.

The efficacy of artificial devices is determined by several factors such as the electro-mechanical characteristics of the device itself, energy transmission characteristics resulting from the coupling of the vibrating source with tissues in the neck region.

---

1 Blom (1979) reports that the most widely known mouth-type electronic larynx consists of a battery-powered pulse generator which connects by a wire to a hand-held tone generator. Sound is directed into the mouth by a short piece of plastic tubing. Similarly, Creech (1976) described a simple and inexpensive modification of a standard Western Electric No. 5 neck type electronic larynx into a mouth-type instrument.
compatibility of the source with vocal tract resonating systems, articulatory dynamics, etc. However, there is a paucity of research which systematically relates these electrical, mechanical, and acoustic variables to speech production with the artificial larynx. Similarly, there is little evidence describing characteristics of the larynx which effect the quality or intelligibility of the speech signal produced with the artificial larynx. The intelligibility of electro-laryngeal speech is dependent upon the physiological, electromechanical and acoustical energies spent in generating an acoustic signal and modifying that signal by respiratory, phonatory and articulatory adjustment.

The development and modification of electrolarynx devices to maximize speech intelligibility must be based upon research which delineates those critical mechanical, acoustic, and physical variables related to the operation of the artificial larynx and total interaction of these devices with the human speech production system.

Our understanding of the critical variables which have had significant effects on the intelligibility of speech signals produced by the electrolarynx requires the evaluation of those variables underpinning the mechanical and acoustic principles of operation of the normal speech production system.

**Acoustic Speech Production**

According to Fant (1960) and Bell, Fujisaki, Heinz, Stevens, and House (1961), a speech wave is the result of acoustic excitation of the vocal tract by one or more sources. All speech sounds,
whether they are vowels or voiced/voiceless consonants, share one common trait. They are produced by a vibratory motion which is sometimes periodic or regular and sometimes aperiodic or turbulent. The vibratory motion of the vibratory source excites the resonating properties of the vocal tract. For normal production of voiced speech sounds the initial source of sound is provided by the vocal cords. This glottal source uses air supplied by the respiratory system to produce a quasi-periodic volume-velocity wave whose spectrum envelope decreases in amplitude with increasing frequency at a rate of about 12dB/octave in the range of 300-2500 Hz. (Bell, Fujisake, Heinz, Stevens, House, 1961). This sound source may produce a variety of fundamental frequencies which are rich in harmonics. The complex tone is introduced into the pharyngeal and oral cavity of the vocal tract and is modified as it interacts with the vocal resonant system. The sound resonating characteristics of the vocal tract are determined at every instant in time by the size and shape of the three-dimensional vocal tube (Rothman, 1978). The quality of the generated sound is also modified by the resonating action of the pharynx, mouth, nose, and by constrictions at the back of the tongue and at the lips and teeth. The different combinations of the shapes of these cavities and the degree of constriction of the tract give rise to various voiced sounds of speech (Barney, 1958). The intensity of the spectrum of the vocal-tract speech output is the sum of a source spectrum, a transfer function related to the filtering process of transmitting the source through the resonating vocal tract, and a radiation characteristic which is the ratio of the sound pressure taken at a distance in
Concentrations of frequency energy with enhanced amplitudes are called formants, and reflect the coupling of harmonic sound energies produced by the vibrating source with the resonant characteristics of the vocal system. These formants are dynamically changing concentrations of frequency energies with varied durations and intensities that are related to the perception of speech (French and Steinberg, 1947).

A set of aerodynamic-mechanical principles similar to those which define speech production of normal speakers must also be applied to the investigation of speech signals produced by laryngectomized patients. In these cases, the laryngeal source of vibration has been removed and the vocal cord tone is completely lost. If an esophageally produced sound cannot be produced, another sound source must be generated. This is the chief function of the artificial larynx. For intelligible elecrolarynx speech, it is essential that the "artificial" tone contain a wide range of harmonics and, for naturalness, the harmonic amplitudes should fall off toward higher frequencies at the same rate as harmonics produced by the real vocal cords (Barney, 1958). The artificial tone should also approximate that of the normal larynx in fundamental frequency variability (Barney, Haworth, Dunn, 1959). Some degree of inflection (frequency variation) is necessary to obtain a considerable improvement of quality. Normal conversational speech contains inflections which correspond to a range of about an octave from the fundamental pitch frequency (Barney, 1958).
Another type of sound which is vital for intelligibility of normal or artificially produced speech is random noise which is produced by the speech apparatus when air passes through a constriction formed by the tongue, teeth or lips. Noise is produced following the release of the articulatory constriction which characterize the production of stops and the production of sibilant consonants. Speech noise may be produced in combination with or in the absence of the vocal-cord tone. Following laryngectomy, a person can no longer produce this random noise in a normal fashion. It is not essential that the artificial larynx produce a substitute noise source for the laryngectomized person. Instead, the air trapped in the mouth and throat can be compressed and expelled to produce the turbulent airflow necessary for noise production. The volume of this alternative noise source is limited, thus limiting the duration and intensity of continuent phoneme productions used during electrolarynx speech (Barney, 1958). The noise which characterizes the production of the /h/ phoneme is usually completely lost when electrolarynx speech is implemented (Barney, 1958).

Comparison of Normally Produced and "Artificially" Produced Source Spectra

The display and analysis of acoustic output of sound sources has been enhanced by the development of various visible speech techniques such as those developed at Bell Telephone Laboratories (Potter, Kopp, and Green, 1947). Through the use of one such technique, sound spectrography, an illustration of the modifications
of harmonic energies of the speech wave can be observed as the vocal tract dimensions change. According to Potter, Kopp, and Green (1947), a narrow band spectrograph will illustrate the individual harmonics of the speech sound. These harmonics appear as closely spaced horizontal lines whose changing intensities determine the over-all harmonic pattern. Sound spectra may be portrayed at any particular time during speech production. The use of such techniques promotes the analysis of formants or groups of harmonics with enhanced amplitudes which reflect sound energies produced by a source coupled with a resonant system. The relationships of formants to one another are distinguishing characteristics of the production and perception of various voiced speech sounds (Peterson and Barney, 1952). Ideally, the sound produced by the artificial larynx will closely approximate the production of natural sounding speech by producing formant patterns resembling those of normal speech sounds.

One variable related to this approximation of normal speech and to the intelligibility of the speech signal is the fundamental frequency of the artificially produced complex tone. The intensity and duration of the artificial signals are also related to speech intelligibility (French and Steinberg, 1947).

Spectrographic analyses of the intensity of artificially produced signals have been completed to determine design criteria and to develop information about the fundamental and harmonic frequencies, sound pressure levels, etc., which promote the production of intelligible and natural sounding speech. Barney, Haworth, and Dunn (1959) measured sound pressure levels of speech produced by subjects who...
have acquired a moderate amount of proficiency with an artificial larynx. Subjects produced a series of vowels while using the electro-larynx. Results indicated that sound pressure levels characteristic of vowel steady states averaged 70-75dB re: 0.0002 microbars at a distance of three feet from the speaker's mouth. This intensity level approximates levels produced during normal speech conversation. These data indicate that the artificial larynx transmits sufficient power into the pharynx throughout the frequency spectrum to permit satisfactory development of the high intensity regions (formants) of the vowel sounds.

It has also been demonstrated by Barney (1958) that the harmonics in the source spectrum of the natural voice are most intense at the low frequencies, dropping in amplitude toward the high frequencies at approximately the inverse 1.5 power of the harmonic number. Spectrographic analysis of the speech signal produced by an artificial larynx displaying a similar pattern of harmonics indicates that it closely approximates the intensity spectra of normal speech (Barney, 1958).

As stated earlier, the output spectra of speech signals also depend upon the critical interactions of the source spectra and the resonant characteristics of the vocal tract. Fant (1960) and others have demonstrated that if the frequency of a periodic disturbance is far removed from the natural resonant frequency of a resonant system, little energy will be transferred from the source to the resonator. Conversely, as the harmonic energies of the source approach the resonant frequencies of the system, an increasingly large amount
of the energy is imparted to the resonant system. The exact form of this transfer function depends on the frictional and volume properties of the system. As resistance of the system increases because of friction, a damping of harmonic energies develops a more broadly tuned system which may be excited by a wider range of harmonic energies. This tuning principle suggests that the fundamental frequency and the harmonic frequencies related to the fundamental are critically related to the spectral characteristics of the speech output signal. It seems critical, then, that the frequency energies produced by the electrolarynx are consonant with vocal tract characteristics for efficient sound production.

In order to more closely examine the frequency energies produced by the electrolarynx implemented in this study, a spectrographic display of vowel steady states produced by a speaker using an electrolarynx at each of the five frequencies was completed. Wide band spectrogram displays of the vowels /i/ and /a/ can be found in Appendices F and G. A narrow band section display of a discrete steady state portion of the vowel was also completed to display the harmonics of the vowels produced by the electrolarynx in an unloaded (not coupled with the vocal tract) condition. In this unloaded condition, complex energies similar to those emitted laryngeally were produced. However, other concentrations of energy were observed at the high frequency end of the spectrum (approximately 8000 Hz.) on the spectrograms for all five of the test frequencies. The source of these high frequency energies was not determined. However, they are not unlike energies which would be emitted when
the metal plunger positioned beneath the plastic diaphragm of the electrolarynx makes contact with the plastic diaphragm surface. Thus, a sound source other than the fundamental and its harmonics is produced. The effect of this source is thought to be minimal because its frequency is far removed from those frequencies considered important for speech intelligibility (0-4000Hz.) and its relative intensity level is insufficient to mask other complex tones. See Appendix H for the sections of each of five fundamental frequencies. Spectrographic data were not further analyzed in the study, but were generated for future analyses.

Perception of Speech Sounds

A distinguishing characteristic of speech is movement. Conversation at the rate of 200 words per minute, corresponding to about four syllables and ten speech sounds per second, is not unusual (French and Steinberg, 1947). During the brief period that a sound lasts, the intensity builds up rapidly, remains comparatively constant for a short time (30-200 msec) then decays rapidly. The various sounds differ from each other in their build-up and decay characteristics in duration, in total intensity, and in the distribution of the intensity with frequency (French and Steinberg, 1947). With the vowel sounds, energy tends to be concentrated in one or more distinct frequency regions, each sound having its own characteristic regions of prominence. The consonant sounds as a group have acoustic energies of higher frequency and lower intensity than the vowel sounds. In addition, the intensity tends to be rather consistent
over the entire frequency regions characteristic of each sound. Thus, when discrete speech sounds are combined in sequence to form syllables, words, and phrases, there is a continuous succession of rapid variations in intensity, not only in particular frequency regions but also along the frequency scale. The interpretation of speech received by the ear depends upon the perception and recognition of these constantly shifting patterns. The intelligibility and quality of the patterns produced depends upon several factors which are discussed below.

Speech Intelligibility

Methods for measuring the intelligibility of speech in general, and of specific speech sounds have been improved and refined steadily since Fletcher and Steinberg reported their data in 1946. The methods have been applied to two directions of research effort, one having to do with the development of effective communications equipment and the other having to do with clinical considerations related to hearing loss specifically, and speech perception more generally (Hirsh, Reynolds, Joseph, 1954). These two important practical situations require that the psychophysical variables related to speech intelligibility, and correlations between those variables and speech contexts characteristic of the many different intelligibility tests be completed.

Intelligibility, as one aspect of the perception of oral communication may be defined as the recognition of speech signals (Hyman, 1979). For a speech signal to be intelligible, both speaker
and listener variables are involved. If adequate distortion or interference of the speech signal occurs during the production of the speech signal by the speaker, by extraneous environmental conditions, or by the perceptual constraints of the listener, an interference in intelligibility and disruption of the oral communication between the speaker and the listener will result. Various aspects of speech production related to articulation, loudness, pitch, rate, quality, auditory and visual cues, noise, perceived acceptability, effectiveness, and quality have been studied to determine possible variables critical to speech intelligibility (Hyman, 1979). Hyman (1979) suggests that intelligible alaryngeal speakers are those for whom the listener can correctly identify at least 90 percent of their articulations, have adequate loudness for the situation, control respiratory air so that it does not mask their speech, and speak at a rate of approximately 120 words per minute. In addition, those alaryngeal speakers that are judged "most acceptable" have pitch levels and inflection variations approximating those of the laryngeal speaker.

While Hyman's speculations may be valid, investigations analyzing the systematic contributions of frequency, intensity or durational characteristics of electrolarynx speech and speech intelligibility are lacking. Rather than emphasizing comparisons of alternative modes of communication as Hyman and others have done, it seems critical and appropriate to investigate the physiological, acoustic, and psychophysical variables which effect speech intelligibility when the electrolarynx is used as a sound generation system.
Speech Quality

There is a dearth of objective information comparing the quality of the laryngectomized individuals who use esophageal speech to those individuals who employ an artificial larynxe (McCroskey and Mulligan, 1963). However, some studies which have emphasized analysis of acoustic variables have also provided qualitative information about speech production with the electrolarynx (Snidecor and Curry, 1959; Bennett and Weinberg, 1973; Crouse, 1962; Snidecor, 1968).

An investigation of frequency, duration, and perceptual measures of alaryngeal speech in relation to judgments of alaryngeal speech acceptability was conducted by Shipp (1967). Results indicated that alaryngeal speech will be rated as more acceptable when the mean fundamental frequency is higher, the duration of the utterance is shorter, the utterance is perceived as having less respiratory noise, and the voice has a higher percentage of measurable phonation with a smaller percentage of silence within the utterance. Reports by Snidecor and Curry (1959) and Curry and Snidecor (1961) also indicated that the mean pitch level of the esophageal speaker is almost exactly one full octave below that of the normal speaker. They also determined that the speaker's ability to demonstrate variable pitch range may have some impact on listeners' acceptance of the speaker as well as intelligibility of the speaker.

Results of studies investigating the perceptual judgments of listeners' preferences for various types of alaryngeal voice have been reported by Bennett and Weinberg (1973). They concluded that
superior esophageal speech was significantly more acceptable than speech produced by highly competent users of the Western Electric reed larynx and the Bell electrolarynx. This finding is in agreement with the work of Crouse (1967) and Snider (1968) which showed that esophageal speech was preferred over speech produced with an electrolarynx, but contradicts Hyman's (1955) conclusion that speech produced with a reed-type artificial larynx is more acceptable than esophageal speech.

Other studies investigating the quality of the artificial sound source have been conducted by Kelly (1979). By manipulating the parameters of intensity, pitch, and on/off control, Kelly reported that a noticeable change in the quality of utterance results. Lack of specificity of procedures or measures taken precludes interpretation or discussion of the methods or results of that study. Investigations conducted by Barney (1958) concluded that for speech to have a natural quality, it should have pitch inflection, both voiced and unvoiced types of output, and frequency/intensity energy spectra which approximate spectra of normally produced speech.

Other Facts Related To the Efficient Production of Electrolarynx Speech

Comparisons of speech quality for different source placement locations indicate that the artificial tone should be applied at the pharynx (Barney, Haworth and Dunn, 1959). Barney (1958) observed that with sound source inserted over the thyroid lamina lateral to the thyroid protuberance speech had a more natural quality than when the source was inserted into the mouth. By using quality estimations of listeners
Barney (1958) concluded that the most natural sounding speech results when the output spectra are shaped using the electrolarynx placed at the laryngeal level.

Since it appears that the preferred position for the sound source is an external placement at the level of the pharynx, it is important to consider the advantages and disadvantages of introducing the sound by placement outside of the throat by means of an artificial larynx. This placement presents two obstacles to the achievement of satisfactory artificial speech. 1) A large amount of power output is required of the vibrator in order to transmit an acoustic signal through the flesh and cartilage into the pharynx. The production of a natural sounding speech output requires a source spectrum which has relatively intense low frequency components. The acoustic problems attendant to efficiently radiating a broad band of frequencies extending from approximately 100 to several thousand cycles with a small transducer are formidable. 2) With an external transducer the sound generated is also directly radiated into the air. This sound which is unmodulated by the vocal tract, produces a competing noise which is perceived as a steady buzz accompanying the filtered speech output signal.

Several other sources of energy loss are present during the production of the acoustic speech signal: a) sound energy radiation from the mouth opening; b) absorption of sound by the walls of the vocal tract; c) loss of sound energy at the glottal opening; d) absorption of sound in the nasal cavity (which has much higher absorption losses than the vocal tract) and; e) loss due to viscosity at the walls of the tract, particularly when narrow constrictions exist in the vocal tract (House and Stevens, 1961).
Statement of the Problem

The aspects of sound production described above indicate that both mechanical and acoustic transmission characteristics of the artificial larynx must be evaluated more completely to maximize the efficiency of these artificial systems. Given that the best tuned acoustic system will yield the best transference of energy at its resonant frequency (Fant, 1960), and based upon the principle that maximum efficiency of sound production is in some way dependent on an ideal match of sound source and resonant frequencies of the vocal tract, it seems reasonable to argue that both the speech spectra and the intelligibility of speech produced by the artificial larynx will vary when different fundamental frequencies and harmonic energies are produced by an artificial sound source. Since any variation of the frequency of the artificial vibrating larynx will alter the source filter characteristics of speech, it seems appropriate to suggest that speech intelligibility will differ significantly within speakers (whose vocal tract resonances are unique) for different frequencies of artificial sound sources and across different speakers. For this experiment it was hypothesized that a) speech intelligibility scores would be significantly different for different frequencies of electrolarynx vibration within speakers and b) intelligibility of different speakers will differ within frequencies. The immediate concerns of this study are a) to delineate frequency differences between and within speakers which might be more thoroughly investigated in future investigations and b) to determine the clinical implications.
which might be derived from these findings. Findings may suggest alteration of the artificial system which allow for adjustments of frequency and intensity characteristics of the vibrating source. Such alterations may enhance the intelligibility of electrolarynx speech for a significant part of the laryngectomized population. Similarities of intelligibility scores between several frequencies within a speaker may promote the use of a fundamental frequency which yields most acceptable quality without significant loss of intelligibility. Additional experimentation may lead to: 1) improvement of the design characteristics of the artificial larynx, 2) improvement of management techniques for training the user of the artificial larynx, 3) the development of additional hypotheses about coupling the artificial larynx to the users neck, damping effects which occur during energy transmission, the articulation patterns of the speaker, and the interaction of these, 4) more important and accurate information pertaining to the role of interaction of vibrating sources and vocal tract in speech intelligibility.
CHAPTER II

METHODS AND PROCEDURES

Speakers

Two adult females and one adult male were selected as speakers for this study. All speakers had normal hearing, normal articulation, and had no previous training with the use of an artificial larynx. Speakers were also selected to represent a continuum of physiological and anatomical differences based upon sex and physique. Assuming that individuals of different stature would differ with regard to vocal tract size and sound transmission characteristics, this selection criterion was used to obviate intelligibility differences which might be related to the interaction of fundamental frequency and differing resonant characteristics of various sized vocal tracts if such differences exist. The ages of speakers ranged from 22 years to 26 years with the mean age being 24.5 years. The two female subjects ($S_1$, $S_2$) were selected from a population of graduate students studying at the Western Michigan University. The third speaker (male) was a student at the Western Theological Seminary.

A Vis-i-Pitch frequency analyzer (Model #4590) was used to compute an average fundamental frequency for each speaker producing the vowel /a/. Since speakers were judged to have normal voice quality, it was assumed that these measures approximated the natural resonant frequency of each speaker. The measured frequencies were 190 Hz. and 230 Hz. for Speakers 1 and 2 respectively, and 100 Hz. for Speaker 3.
Stimulus Materials

Five lists of PB-50 words (CVC type) were randomly chosen from the USA Standard Monosyllabic Word Intelligibility List (see Appendix A). Each PB word was embedded in the sentence "You will say _____". The carrier phrase was implemented to a) minimize context effects on the intelligibility of PB words and b) provide an utterance sequence to stabilize vocal intensity levels prior to and during the production of each stimulus word.

Familiarity of words was taken under consideration during the initial stages of selection. A total of 15 word lists were reviewed. A list was eliminated if it contained two or more words that were unfamiliar to the examiner. Five lists were then randomly selected from those remaining. It was assumed that this procedure would minimize the probability of intelligibility errors due to unfamiliarity.

Apparatus

A Western Electric No. 5 electric larynx was used as a vibratory source. This source was modified to produce five arbitrary fundamental frequencies of vibration (94 Hz., 120 Hz., 160 Hz., 212 Hz., and 245 Hz.) by introducing a modified switch into the electrolarynx circuitry (see Appendix B). A frequency meter (Hewlett Packard, Model 5300) was used to determine the frequency settings. The electrolarynx was driven by a constant voltage power supply (Heathkit IP-27) to minimize variability of frequency and intensity of the sound source.
Sound pressure levels of the various frequencies of the electrolarynx under loaded and unloaded conditions were determined in order to observe intensity changes resulting from the coupling of the electrolarynx to the user's neck. In the unloaded condition, the electrolarynx was held three feet from a sound level meter with the vibrating diaphragm facing the microphone. For the loaded condition, the electrolarynx was placed posteriolaterally with reference to the laryngeal protuberance. The microphone of the sound level meter (B&K #4144) was placed 3 feet from the diaphragm of the electrolarynx. Two "loaded" sound pressure level values were measured to determine intensities when a) the vocal cavity was occluded (mouth closed; tongue apex and tip in contact with the maxillary surface) and b) the vocal tract was open in a position approximating that assumed for the production of the vowel /a/. The intensity of the signal decreased by approximately 12dB when coupled with the "open" vocal tract and by approximately 21dB when coupled with the "closed" tract. However, intensity changes were generally consistent across frequencies. An intensity difference of 5dB between frequencies 212 Hz. and 245 Hz. was observed for the "open" condition. Similarly, a difference of 6dB between frequencies 94 Hz. and 245 Hz. for the "closed" condition was observed. This difference observed in the "closed" condition may be related to resonant characteristics of the speakers vocal tract which may either enhance the transfer of high frequency energies (245 Hz.) or depress the transfer of low frequency (94 Hz.) energies. Intensity measures are listed in Appendix C.
Recording and Presentation of Stimulus Words

Recording

A dynamic microphone (Shure 570) was connected to a head piece and positioned 1" from the corner of the speaker's lips at a 45 degree angle to a midsagittal reference plane.

Three speakers were selected to make recordings of 5 lists of CVC word lists using a Western Electric No. 5 electric larynx. A brief training period for the speakers was conducted to instruct them with the operation of the electrolarynx. All speakers were given approximately the same amount of time (15 minutes) to practice using the artificial larynx.

Following the training segment, each speaker was instructed to produce each word on the list by attaching it to the carrier phrase "You will say ______." Speakers were instructed and trained to produce the entire utterance at a constant intensity level by peaking a VU meter at a constant reference level for each carrier phrase and test word produced. Consistent within-subject intensity levels were maintained across frequencies and word lists.

Word lists were recorded on an Ampex (AG-350) tape recorder in a sound treated recording room. Speakers were instructed to maintain an adductory vocal cord position for all recordings to eliminate the use of pulmonary air during recordings. A custom made blinking light timer was used to control the speaker's rate. A two second interval separated each sentence production. Each speaker produced the 50 word list for each test frequency pausing between recordings for
each frequency. When each list had been recorded, the speaker was
given a 5 minute rest and the next list was recorded, at a different
fundamental frequency. Prior to the recording session, lists for each
frequency were selected and ordered according to a scheme (Appendix
B) which minimized the introduction of an order effect: due to the
experimental procedure.

Presentation

Three listening groups of 7 listeners were selected from a
population of graduate and undergraduate students studying at
Western Michigan University. Because of time and scheduling
limitations, two of the listening groups participated during the final
week of their semester studies. The other group participated at
mid-semester.

Each listener was presented with five sets of 50 CVC word
lists for each of three speakers at each of five frequencies.
Speakers and frequencies were presented in different orders for
each listening group to minimize effects of sequencing or practice
(see Appendix E). An Ampex Tape recorder/reproducer (AG-350),
a Grason-Stadler speech audiometer (162), and a JVC Audio Amplifier
(JA-555) were used to reproduce the recorded utterances in a sound
treated room. The recordings were presented through a loudspeaker
(Electra-voice, Model SP-12) at a listening level of approximately
70dB SPL at the listener's ears. Listeners were provided 2 seconds
to make each response. This latency was considered to be sufficient based
upon pilot data.
Following the intelligibility portion of the experiment, listeners were instructed to estimate the sex of each speaker and to rate each speaker on a scale of 1-7 for acceptability (Appendix J).
CHAPTER III

RESULTS

Introduction

Three speakers (2 female, 1 male) produced and recorded monosyllabic word lists using a Western Electric No. 5 artificial larynx which had been adjusted to produce five frequencies of vibration (94 Hz., 120 Hz., 160 Hz., 212 Hz., 245 Hz.). The twenty-one listeners recorded the perceived words on answer sheets (see Appendix L). A percentage of correct responses for each word list at each frequency for each speaker was subsequently determined (Table 1). Mean intelligibility scores were calculated for each listening group (Table 2). Statistical analyses were implemented to determine the relationship of listeners' intelligibility scores to the frequency mode of vibration and the differences of these relationships within and across speakers.

Intelligibility Scores

Intelligibility scores for each listener were determined by calculating percentages of correct responses for each speaker at each of the five test frequencies and for each of three groups of listeners (see Table 1). Mean scores were computed for each of the three speakers by frequency and by group.

Statistical Procedures

A multivariate analysis of variance program for repeated
Table 1
Intelligibility Scores (Percentages) for Each Speaker At Each of Five Test Frequencies (A=94Hz., B=120Hz., C=160Hz., D=212Hz.
E=245Hz.) For Each Listener For Three Groups of Listeners

<table>
<thead>
<tr>
<th>Frequency</th>
<th>Speaker 1</th>
<th>Speaker 2</th>
<th>Speaker 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
</tr>
<tr>
<td>Group 1</td>
<td>Listener 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>82 82 86 82 84</td>
<td>60 50 40 44 66</td>
<td>48 34 42 34 40</td>
</tr>
<tr>
<td>2</td>
<td>54 64 62 64 66</td>
<td>32 42 40 22 44</td>
<td>30 24 32 20 16</td>
</tr>
<tr>
<td>3</td>
<td>78 80 84 86 84</td>
<td>64 56 42 42 58</td>
<td>50 26 56 40 46</td>
</tr>
<tr>
<td>4</td>
<td>86 82 84 90 82</td>
<td>54 46 50 52 52</td>
<td>58 32 48 42 30</td>
</tr>
<tr>
<td>5</td>
<td>62 76 76 78 76</td>
<td>58 54 42 48 60</td>
<td>44 32 40 44 64</td>
</tr>
<tr>
<td>6</td>
<td>72 70 70 70 70</td>
<td>38 46 28 38 46</td>
<td>48 14 26 24 14</td>
</tr>
<tr>
<td>7</td>
<td>70 68 80 86 70</td>
<td>44 58 36 46 72</td>
<td>46 30 52 42 28</td>
</tr>
<tr>
<td>Group 2</td>
<td>Listener 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>52 56 56 58 54</td>
<td>34 40 30 26 48</td>
<td>22 18 36 20 12</td>
</tr>
<tr>
<td>2</td>
<td>58 60 58 76 68</td>
<td>50 52 20 28 48</td>
<td>38 22 50 38 40</td>
</tr>
<tr>
<td>3</td>
<td>70 80 78 80 82</td>
<td>62 54 42 30 58</td>
<td>46 34 46 44 40</td>
</tr>
<tr>
<td>4</td>
<td>70 88 70 74 64</td>
<td>46 54 46 36 58</td>
<td>50 42 52 40 42</td>
</tr>
<tr>
<td>5</td>
<td>76 82 70 76 66</td>
<td>64 56 50 36 64</td>
<td>70 32 58 36 46</td>
</tr>
<tr>
<td>6</td>
<td>68 80 68 80 82</td>
<td>66 64 38 44 58</td>
<td>44 42 64 42 38</td>
</tr>
<tr>
<td>7</td>
<td>58 70 60 64 68</td>
<td>60 36 28 42 52</td>
<td>28 36 46 38 42</td>
</tr>
<tr>
<td>Group 3</td>
<td>Listener 3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>56 66 56 62 58</td>
<td>48 42 38 42 54</td>
<td>36 26 38 32 32</td>
</tr>
<tr>
<td>2</td>
<td>50 64 68 68 72</td>
<td>50 54 28 46 62</td>
<td>42 26 50 40 42</td>
</tr>
<tr>
<td>3</td>
<td>42 58 60 64 66</td>
<td>52 48 44 46 56</td>
<td>34 26 34 32 24</td>
</tr>
<tr>
<td>4</td>
<td>52 64 62 70 64</td>
<td>52 50 48 54 74</td>
<td>40 34 52 28 48</td>
</tr>
<tr>
<td>5</td>
<td>58 62 66 72 72</td>
<td>56 50 40 52 70</td>
<td>48 18 44 28 42</td>
</tr>
<tr>
<td>6</td>
<td>52 68 64 70 72</td>
<td>56 48 40 52 64</td>
<td>34 26 48 42 30</td>
</tr>
<tr>
<td>7</td>
<td>46 54 40 50 44</td>
<td>48 40 36 36 62</td>
<td>40 16 28 20 24</td>
</tr>
</tbody>
</table>
Table 2

Individual and Pool Group Mean Intelligibility Scores for Each Speaker Across Frequencies

<table>
<thead>
<tr>
<th>Frequency</th>
<th>Speaker 1</th>
<th>Speaker 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>94Hz.</td>
<td>120Hz.</td>
</tr>
<tr>
<td>Group 1</td>
<td>72.00</td>
<td>74.57</td>
</tr>
<tr>
<td>Group 2</td>
<td>64.57</td>
<td>73.71</td>
</tr>
<tr>
<td>Group 3</td>
<td>50.86</td>
<td>62.28</td>
</tr>
<tr>
<td>Pooled</td>
<td>62.48</td>
<td>70.19</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Frequency</th>
<th>Speaker 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>94Hz.</td>
</tr>
<tr>
<td>Group 1</td>
<td>46.29</td>
</tr>
<tr>
<td>Group 2</td>
<td>42.57</td>
</tr>
<tr>
<td>Group 3</td>
<td>39.14</td>
</tr>
<tr>
<td>Pooled</td>
<td>42.67</td>
</tr>
</tbody>
</table>
measures (Western Michigan University Department of Mathematics, 1982) was executed to compare speakers across listening groups, frequencies across speakers, and frequencies across groups.

Correlational t tests were calculated to determine if significant differences of intelligibility scores existed

a) between the three speakers for each frequency,

b) between the test frequencies within each speaker, and

c) between the three listening groups for speakers and frequencies. (Table 3). All statements of significance for comparisons are with reference to the .05 probability level.

The Relationship Between Fundamental Frequency (F₀) Of Electrolarynx Vibration And Speaker Intelligibility Scores For Speakers

Interaction Between Listening Groups and Speakers

An interaction of listening groups, speakers and frequencies (p=.026) was observed when statistical analyses were completed to determine if the responses of the 3 groups were uniform. The interaction indicated that scores of listening groups were different for different frequencies and speakers. However, a strong relationship between frequency of vibration and intelligibility scores within speakers (Table 4) was evident despite the existence of this interaction. The significance of the group interaction effect indicated that pooling the 3 listening groups for further statistical comparison might reduce the probability of observing statistically significant frequency differences if they exist. Therefore, it was determined that data would be analyzed for each listening group as well as for pooled data. The strong trends observed between the frequency of vibration and
Table 3

Pooled Mean Intelligibility Scores Across Frequencies
For Speaker 1 (S₁), Speaker 2 (S₂), and Speaker 3 (S₃)

<table>
<thead>
<tr>
<th></th>
<th>A(94Hz)</th>
<th>B(120Hz)</th>
<th>C(160Hz)</th>
<th>D(212Hz)</th>
<th>E(245Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>62</td>
<td>52</td>
<td>42</td>
<td>49</td>
<td>38</td>
</tr>
<tr>
<td>S2</td>
<td>70</td>
<td>49</td>
<td>28</td>
<td>38</td>
<td>41</td>
</tr>
<tr>
<td>S3</td>
<td>67</td>
<td>44</td>
<td>44</td>
<td>41</td>
<td>34</td>
</tr>
<tr>
<td>S1</td>
<td>72</td>
<td>72</td>
<td>72</td>
<td>72</td>
<td>72</td>
</tr>
<tr>
<td>S2</td>
<td>68</td>
<td>68</td>
<td>68</td>
<td>68</td>
<td>68</td>
</tr>
<tr>
<td>S3</td>
<td>58</td>
<td>58</td>
<td>58</td>
<td>58</td>
<td>58</td>
</tr>
<tr>
<td>S1</td>
<td>35</td>
<td>35</td>
<td>35</td>
<td>35</td>
<td>35</td>
</tr>
<tr>
<td>S2</td>
<td>35</td>
<td>35</td>
<td>35</td>
<td>35</td>
<td>35</td>
</tr>
<tr>
<td>S3</td>
<td>35</td>
<td>35</td>
<td>35</td>
<td>35</td>
<td>35</td>
</tr>
</tbody>
</table>
Table 4

Rank Order of Intelligibility (Low-High) of Test Frequencies For Listening Groups. Frequency A=94Hz., Frequency B=120Hz., Frequency C=160Hz., Frequency D=212Hz., Frequency E=245Hz. Lines Below Frequency Orders Depict Intelligibility Scores for Frequencies Which Were Not Significantly Different (p .05)

<table>
<thead>
<tr>
<th>Group</th>
<th>Speaker 1</th>
<th>Speaker 2</th>
<th>Speaker 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>AEBCD</td>
<td>CDAEB</td>
<td>BEDCA</td>
</tr>
<tr>
<td>2</td>
<td>ACEDB</td>
<td>DCBAE</td>
<td>BDEAC</td>
</tr>
<tr>
<td>3</td>
<td>ACBED</td>
<td>CDBAE</td>
<td>BDEAC</td>
</tr>
<tr>
<td>Pooled</td>
<td>ACEBD</td>
<td>CDBAE</td>
<td>BDEAC</td>
</tr>
</tbody>
</table>

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
intelligibility scores across listening groups also indicated that analysis of pooled data was appropriate. Analysis of pooled data is presented in Table 4 and Appendix I for comparison with findings for individual listening groups.

Intelligibility Comparisons Between Speakers

Comparative analysis of frequencies between speakers was implemented by calculating t probabilities for each speaker for each individual frequency. Mean group listening scores were pooled for this analysis. The t probabilities were used to determine significant differences between speakers for the test frequencies. All t probabilities for interspeaker comparisons by frequency were different from each other at each of the 5 test frequencies.

Results of this analysis also indicated that at all frequencies Speaker 1 was significantly more intelligible than Speakers 2 and 3. Speaker 2 was also significantly more intelligible than Speaker 3 in all cases but for Frequency C (160 Hz.). In that case Speaker 3 was significantly more intelligible than Speaker 2 (Table 3).

Variability of Intelligibility Scores

Mean intelligibility scores for Speaker 1 were more variable than scores for Speakers 2 and 3 at all frequency settings except for 245 Hz. At this frequency scores for Speaker 3 were more variable. Scores for Speaker 2 were more variable than those of Speaker 3 at 94 Hz. and 212 Hz., while scores for Speaker 3 were more variable than those of Speaker 2 at 120 Hz. and 160 Hz.
Intelligibility Scores Within Speakers

Mean intelligibility scores for each speaker were determined within frequencies for individual and pooled listener groups. Scores are presented in Table 5 (Speaker 1), Table 6 (Speaker 2), and Table 7 (Speaker 3).

Speaker 1

Mean intelligibility scores and variability of those scores for Speaker 1 at the test frequencies are reported in Table 5. Results indicated that when the groups were pooled, Speaker 1 was most intelligible at 212 Hz and 120 Hz. Lowest intelligibility scores were obtained at 94 Hz. When compared to the pooled group scores, mean scores for individual listening groups also indicated that 212 Hz. and 120 Hz. yielded the highest intelligibility scores for 2 of the 3 groups and 2nd highest scores for the 3rd listening group (Appendix I). Frequency A (94 Hz.) received the lowest intelligibility scores for Speaker 1 in each listening group.

Speaker 2

Pooled results for Speaker 2 indicated that the highest intelligibility scores were obtained at 245 Hz. Lowest intelligibility scores were obtained at 160 Hz. Mean intelligibility scores for individual groups indicated that each listening group judged 245 Hz. to be the most intelligible (Appendix I). Ranges of scores differed across groups. Similarly, 160 Hz. yielded lowest intelligibility scores
Table 5
Mean Intelligibility Scores and Standard Deviations for Speaker 1 Across Frequencies and Groups

<table>
<thead>
<tr>
<th>Frequency</th>
<th>A(94Hz.)</th>
<th>B(120Hz.)</th>
<th>C(160Hz.)</th>
<th>D(212Hz.)</th>
<th>E(245Hz.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Listener</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>82%</td>
<td>82%</td>
<td>85%</td>
<td>82%</td>
<td>84%</td>
</tr>
<tr>
<td>2</td>
<td>54%</td>
<td>64%</td>
<td>62%</td>
<td>64%</td>
<td>66%</td>
</tr>
<tr>
<td>3</td>
<td>78%</td>
<td>80%</td>
<td>84%</td>
<td>86%</td>
<td>84%</td>
</tr>
<tr>
<td>4</td>
<td>86%</td>
<td>82%</td>
<td>84%</td>
<td>90%</td>
<td>82%</td>
</tr>
<tr>
<td>5</td>
<td>62%</td>
<td>76%</td>
<td>76%</td>
<td>78%</td>
<td>76%</td>
</tr>
<tr>
<td>6</td>
<td>72%</td>
<td>70%</td>
<td>70%</td>
<td>70%</td>
<td>52%</td>
</tr>
<tr>
<td>7</td>
<td>70%</td>
<td>68%</td>
<td>80%</td>
<td>86%</td>
<td>70%</td>
</tr>
<tr>
<td>Group 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>56%</td>
<td>56%</td>
<td>56%</td>
<td>58%</td>
<td>54%</td>
</tr>
<tr>
<td>2</td>
<td>58%</td>
<td>60%</td>
<td>58%</td>
<td>76%</td>
<td>68%</td>
</tr>
<tr>
<td>3</td>
<td>70%</td>
<td>80%</td>
<td>78%</td>
<td>80%</td>
<td>82%</td>
</tr>
<tr>
<td>4</td>
<td>70%</td>
<td>88%</td>
<td>70%</td>
<td>74%</td>
<td>64%</td>
</tr>
<tr>
<td>5</td>
<td>76%</td>
<td>82%</td>
<td>70%</td>
<td>76%</td>
<td>66%</td>
</tr>
<tr>
<td>6</td>
<td>68%</td>
<td>80%</td>
<td>68%</td>
<td>80%</td>
<td>82%</td>
</tr>
<tr>
<td>7</td>
<td>58%</td>
<td>70%</td>
<td>60%</td>
<td>64%</td>
<td>68%</td>
</tr>
<tr>
<td>Group 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>56%</td>
<td>66%</td>
<td>56%</td>
<td>62%</td>
<td>58%</td>
</tr>
<tr>
<td>2</td>
<td>50%</td>
<td>64%</td>
<td>68%</td>
<td>68%</td>
<td>72%</td>
</tr>
<tr>
<td>3</td>
<td>42%</td>
<td>58%</td>
<td>60%</td>
<td>64%</td>
<td>66%</td>
</tr>
<tr>
<td>4</td>
<td>52%</td>
<td>64%</td>
<td>62%</td>
<td>70%</td>
<td>64%</td>
</tr>
<tr>
<td>5</td>
<td>58%</td>
<td>62%</td>
<td>66%</td>
<td>72%</td>
<td>72%</td>
</tr>
<tr>
<td>6</td>
<td>52%</td>
<td>68%</td>
<td>64%</td>
<td>70%</td>
<td>72%</td>
</tr>
<tr>
<td>7</td>
<td>46%</td>
<td>54%</td>
<td>40%</td>
<td>50%</td>
<td>44%</td>
</tr>
<tr>
<td>Group 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\bar{x})</td>
<td>62.48%</td>
<td>70.19%</td>
<td>67.52%</td>
<td>72.38%</td>
<td>68.86%</td>
</tr>
<tr>
<td>(\sigma)</td>
<td>12.24</td>
<td>9.96</td>
<td>11.27</td>
<td>10.03</td>
<td>10.89</td>
</tr>
</tbody>
</table>
Table 6

Mean Intelligibility Scores and Standard Deviations for Speaker 2 Across Frequencies and Groups

<table>
<thead>
<tr>
<th>Frequency</th>
<th>A(94Hz.)</th>
<th>B(120Hz.)</th>
<th>C(160Hz.)</th>
<th>D(212Hz.)</th>
<th>E(245Hz.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Listener</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>60%</td>
<td>50%</td>
<td>40%</td>
<td>44%</td>
<td>66%</td>
</tr>
<tr>
<td>2</td>
<td>32%</td>
<td>42%</td>
<td>40%</td>
<td>22%</td>
<td>33%</td>
</tr>
<tr>
<td>3</td>
<td>64%</td>
<td>56%</td>
<td>42%</td>
<td>42%</td>
<td>58%</td>
</tr>
<tr>
<td>Group 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>54%</td>
<td>46%</td>
<td>50%</td>
<td>52%</td>
<td>52%</td>
</tr>
<tr>
<td>5</td>
<td>58%</td>
<td>54%</td>
<td>42%</td>
<td>48%</td>
<td>60%</td>
</tr>
<tr>
<td>6</td>
<td>38%</td>
<td>46%</td>
<td>28%</td>
<td>38%</td>
<td>46%</td>
</tr>
<tr>
<td>7</td>
<td>44%</td>
<td>58%</td>
<td>36%</td>
<td>46%</td>
<td>72%</td>
</tr>
<tr>
<td>Group 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>34%</td>
<td>40%</td>
<td>30%</td>
<td>26%</td>
<td>48%</td>
</tr>
<tr>
<td>2</td>
<td>50%</td>
<td>52%</td>
<td>20%</td>
<td>28%</td>
<td>48%</td>
</tr>
<tr>
<td>3</td>
<td>62%</td>
<td>54%</td>
<td>42%</td>
<td>30%</td>
<td>58%</td>
</tr>
<tr>
<td>Group 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>46%</td>
<td>54%</td>
<td>46%</td>
<td>36%</td>
<td>58%</td>
</tr>
<tr>
<td>5</td>
<td>64%</td>
<td>56%</td>
<td>50%</td>
<td>36%</td>
<td>64%</td>
</tr>
<tr>
<td>6</td>
<td>66%</td>
<td>64%</td>
<td>38%</td>
<td>44%</td>
<td>58%</td>
</tr>
<tr>
<td>7</td>
<td>60%</td>
<td>46%</td>
<td>28%</td>
<td>42%</td>
<td>52%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>(\bar{x})</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Group</td>
<td>52.10%</td>
<td>49.52%</td>
<td>38.38%</td>
<td>41.04%</td>
<td>58.38%</td>
</tr>
<tr>
<td></td>
<td>9.60</td>
<td>6.92</td>
<td>7.89</td>
<td>9.05</td>
<td>8.36</td>
</tr>
</tbody>
</table>
Table 7

Mean Intelligibility Scores and Standard Deviations for Speaker 3 Across Frequencies and Groups

<table>
<thead>
<tr>
<th>Frequency</th>
<th>A(94Hz.)</th>
<th>B(120Hz.)</th>
<th>C(160Hz.)</th>
<th>D(212Hz.)</th>
<th>E(245Hz.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Listener</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>48%</td>
<td>34%</td>
<td>42%</td>
<td>34%</td>
<td>40%</td>
</tr>
<tr>
<td>2</td>
<td>30%</td>
<td>24%</td>
<td>32%</td>
<td>20%</td>
<td>16%</td>
</tr>
<tr>
<td>3</td>
<td>50%</td>
<td>26%</td>
<td>56%</td>
<td>40%</td>
<td>46%</td>
</tr>
<tr>
<td>4</td>
<td>58%</td>
<td>32%</td>
<td>48%</td>
<td>42%</td>
<td>30%</td>
</tr>
<tr>
<td>5</td>
<td>44%</td>
<td>32%</td>
<td>40%</td>
<td>44%</td>
<td>64%</td>
</tr>
<tr>
<td>6</td>
<td>48%</td>
<td>14%</td>
<td>26%</td>
<td>24%</td>
<td>14%</td>
</tr>
<tr>
<td>7</td>
<td>46%</td>
<td>30%</td>
<td>52%</td>
<td>42%</td>
<td>28%</td>
</tr>
<tr>
<td>Group 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>20%</td>
<td>18%</td>
<td>36%</td>
<td>20%</td>
<td>12%</td>
</tr>
<tr>
<td>2</td>
<td>38%</td>
<td>22%</td>
<td>50%</td>
<td>38%</td>
<td>40%</td>
</tr>
<tr>
<td>3</td>
<td>46%</td>
<td>34%</td>
<td>46%</td>
<td>44%</td>
<td>40%</td>
</tr>
<tr>
<td>4</td>
<td>50%</td>
<td>42%</td>
<td>52%</td>
<td>40%</td>
<td>42%</td>
</tr>
<tr>
<td>5</td>
<td>70%</td>
<td>32%</td>
<td>58%</td>
<td>36%</td>
<td>46%</td>
</tr>
<tr>
<td>6</td>
<td>44%</td>
<td>42%</td>
<td>64%</td>
<td>42%</td>
<td>38%</td>
</tr>
<tr>
<td>7</td>
<td>28%</td>
<td>36%</td>
<td>46%</td>
<td>38%</td>
<td>42%</td>
</tr>
<tr>
<td>Group 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>36%</td>
<td>26%</td>
<td>38%</td>
<td>32%</td>
<td>32%</td>
</tr>
<tr>
<td>2</td>
<td>42%</td>
<td>26%</td>
<td>50%</td>
<td>40%</td>
<td>40%</td>
</tr>
<tr>
<td>3</td>
<td>34%</td>
<td>26%</td>
<td>34%</td>
<td>32%</td>
<td>28%</td>
</tr>
<tr>
<td>4</td>
<td>40%</td>
<td>34%</td>
<td>52%</td>
<td>28%</td>
<td>48%</td>
</tr>
<tr>
<td>5</td>
<td>48%</td>
<td>18%</td>
<td>44%</td>
<td>28%</td>
<td>42%</td>
</tr>
<tr>
<td>6</td>
<td>34%</td>
<td>26%</td>
<td>48%</td>
<td>42%</td>
<td>30%</td>
</tr>
<tr>
<td>7</td>
<td>40%</td>
<td>16%</td>
<td>28%</td>
<td>20%</td>
<td>24%</td>
</tr>
<tr>
<td>Group 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[
\bar{x} = 42.67\% \quad 28.98\% \quad 44.86\% \quad 34.57\% \quad 35.43\%
\]

\[
\sigma = 10.57 \quad 7.84 \quad 9.93 \quad 8.25 \quad 12.52
\]
for 2 out of 3 groups and 2nd to the lowest in the 3rd listening group. Intelligibility scores at 94 Hz. were second highest for the pooled data and for 2 out of 3 individual listening groups.

**Speaker 3**

Pooled results for Speaker 3 indicated that the highest intelligibility scores for Speaker 3 were obtained at 160 Hz. Lowest intelligibility scores were obtained at 120 Hz. In comparison to the pooled group scores, the individual group mean scores indicated that 160 Hz. yielded the highest intelligibility scores 2 out of 3 groups and 2nd highest in the 3rd group. Frequency A (94 Hz.) intelligibility scores were 2nd highest for 2 out of 3 groups and highest for the 3rd group (Appendix I).

Summary: The Relationship Between Fundamental Frequency And Speech Intelligibility Scores Within And Between Speakers

Results of Correlational t test comparisons indicated that for each test frequency there were significant differences of intelligibility scores between speakers. Results for listening groups and pooled data indicated that Speaker 1 was significantly more intelligible than Speakers 2 and 3 and Speaker 2 was significantly more intelligible than Speaker 3.

Differences between intelligibility scores obtained for five test frequencies within a given speaker were also observed. Maximum intelligibility scores were more highly correlated with certain frequencies for certain speakers (Table 8). Results indicated that
the relation of maximum intelligibility scores and fundamental frequency of vibration is speaker dependent. Data illustrated in Appendix I indicated that Speaker 1 obtained the highest intelligibility scores when the electrolarynx was set at 212 Hz. and 120 Hz. and the lowest intelligibility scores at 94 Hz. Speaker 2 obtained significantly higher intelligibility scores at 245 Hz. and significantly lower scores at 160 Hz. and 212 Hz. Speaker 3 obtained the highest intelligibility scores at 160 Hz. and 94 Hz. and significantly lower scores at 120 Hz. for both the pooled and individual data.

It may be concluded that some frequencies yield significantly higher intelligibility scores for certain speakers. Given that dynamic interaction of the electrical, mechanical, and acoustic aspects of the vibrating electrolarynx and the speaker's vocal tract are different across speakers, the observed intelligibility scores are not surprising. However, they indicate that the coupling of an individual's sound resonating cavity with the electrolarynx, and the fundamental frequency of vibration of that electrolarynx are systematically related to intelligibility of the speech signal.

Analysis of Phoneme Errors By Frequency

While the comparisons of intelligibility scores to various fundamental frequencies of electrolarynx vibration supported the hypothesis that certain frequencies of vibration will be more intelligible for different speakers, the variables underpinning these differences have not been clarified. As a means of determining possible sources of these differences, a description of phoneme errors which characterized intelligibility differences for each
Table 8
Pooled Group Mean Intelligibility Scores for Speakers at Each Frequency

<table>
<thead>
<tr>
<th>Frequency</th>
<th>Speaker 1</th>
<th>Speaker 2</th>
<th>Speaker 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>A (94Hz.)</td>
<td>62.48</td>
<td>52.09</td>
<td>42.66</td>
</tr>
<tr>
<td>B (120Hz.)</td>
<td>70.19</td>
<td>49.52</td>
<td>28.09</td>
</tr>
<tr>
<td>C (160 Hz.)</td>
<td>61.19</td>
<td>38.38</td>
<td>44.85</td>
</tr>
<tr>
<td>D (212Hz.)</td>
<td>72.38</td>
<td>41.04</td>
<td>34.57</td>
</tr>
<tr>
<td>E (245Hz.)</td>
<td>68.86</td>
<td>58.38</td>
<td>35.43</td>
</tr>
</tbody>
</table>
frequency were completed. No consistent patterns of frequency vs. phoneme error types were observed. However, certain Speaker dependent patterns were observed. A description of error types by Speaker and frequency is provided below.

A series of matrices (Appendix M) was developed to determine the percentage of phoneme error types made by each Speaker at each frequency. Errors of manner (plosive, fricative, affricate, nasal, liquid, glide) and place (bilabial, labiodental, lingualalveolar, and velar) were determined by analyzing a random selection of 3 written response packets from each group. Trends were so evident that analysis of all 21 listener's responses at all frequencies was not seen to be necessary in order to obtain a representative sample for purposes of this study. The percentage of phoneme errors by frequency and speaker are shown in Tables 9, 10, 11 and Appendix M. Specific error types were not analyzed statistically for this experiment. However, data are provided for discussion.

Speaker 1

Substitution errors characteristic of Speaker 1 consisted of voicing substitution errors of the plosives b/p (20%), d/t (23%), and g/k (18%) and the fricatives z/s and dʃ/ʃ (11%). Omission of the phoneme /h/ was observed to comprise 8% of the error total. The nasals /m/, and /n/ were frequently substituted by the voiced phoneme /b/ (1.6%) while the voiceless phoneme /p/ was often substituted by the voiced nasal phonemes /m/ and /n/ (2%). Phoneme substitution errors did not appear to be systematically
Table 9

Matrix of Total (577) Phoneme Substitution Errors and Percentage of Total Phoneme Errors (Parenthesized) for Speaker 1 Across All Frequencies and Listeners. Only Substitution Types Accounting For More Than 1% of the Error Total are Presented

<table>
<thead>
<tr>
<th></th>
<th>p</th>
<th>t</th>
<th>k</th>
<th>b</th>
<th>d</th>
<th>g</th>
<th>glide</th>
<th>fric.</th>
<th>omis.</th>
<th>m/n</th>
<th>vowel</th>
</tr>
</thead>
<tbody>
<tr>
<td>p</td>
<td></td>
<td>116</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5</td>
<td>(.009)</td>
<td></td>
<td></td>
<td>5</td>
</tr>
<tr>
<td>t</td>
<td></td>
<td></td>
<td>135</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(.233)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>k</td>
<td></td>
<td></td>
<td></td>
<td>104</td>
<td></td>
<td></td>
<td></td>
<td>(.180)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>12</td>
<td>(.020)</td>
<td></td>
</tr>
<tr>
<td>d</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>g</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>8</td>
<td>(.014)</td>
</tr>
<tr>
<td>glide</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>fric.</td>
<td></td>
<td></td>
<td></td>
<td>46</td>
<td></td>
<td></td>
<td></td>
<td>(.078)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>omis.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>67</td>
<td>(.116)</td>
<td></td>
</tr>
<tr>
<td>m/n</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>9</td>
<td>(.016)</td>
<td></td>
</tr>
<tr>
<td>h</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>50</td>
<td>(.087)</td>
</tr>
<tr>
<td>r</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>added</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>14</td>
<td>(.024)</td>
</tr>
<tr>
<td>vowel</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>59</td>
</tr>
</tbody>
</table>
Table 10

Matrix of Total (851) Phoneme Substitution Errors and Percentage of Total Phoneme Errors (Parenthesized) for Speaker 2 Across All Frequencies and Listeners. Only Substitution Types Accounting for More Than 1% of the Error Total Are Presented

<table>
<thead>
<tr>
<th></th>
<th>p</th>
<th>t</th>
<th>k</th>
<th>b</th>
<th>d</th>
<th>g</th>
<th>glide</th>
<th>fric.</th>
<th>omis.</th>
<th>m/n</th>
<th>vowel</th>
</tr>
</thead>
<tbody>
<tr>
<td>p</td>
<td>5</td>
<td>(0.005)</td>
<td>97</td>
<td>(0.114)</td>
<td></td>
<td></td>
<td></td>
<td>49</td>
<td>(0.058)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>t</td>
<td>80</td>
<td>(0.094)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>12</td>
<td>(0.014)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>k</td>
<td>19</td>
<td>(0.022)</td>
<td>9</td>
<td>(0.010)</td>
<td>59</td>
<td>(0.069)</td>
<td></td>
<td>13</td>
<td>(0.015)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>b</td>
<td>8</td>
<td>(0.009)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d</td>
<td>16</td>
<td>(0.019)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>g</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>glide</td>
<td>27</td>
<td>(0.031)</td>
<td>31</td>
<td>(0.036)</td>
<td>70</td>
<td>(0.082)</td>
<td>59</td>
<td>(0.069)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>fric.</td>
<td>15</td>
<td>(0.017)</td>
<td>129</td>
<td>(0.151)</td>
<td>16</td>
<td>(0.019)</td>
<td>48</td>
<td>(0.001)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>omis.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>m/n</td>
<td>11</td>
<td>(0.013)</td>
<td>12</td>
<td>(0.014)</td>
<td></td>
<td></td>
<td></td>
<td>24</td>
<td>(0.028)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>h</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>39</td>
<td>(0.460)</td>
</tr>
<tr>
<td>r</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>25</td>
<td>(0.029)</td>
</tr>
<tr>
<td>added</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>18</td>
<td>(0.021)</td>
</tr>
<tr>
<td>vowel</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>225</td>
</tr>
</tbody>
</table>
Table 11
Matrix of Total (1109) Phoneme Substitution Errors and Percentages of Total Phoneme Errors (Parenthesized) For Speaker 3 Across All Frequencies and Listeners. Only Substitution Types Accounting for 1% of the Error Total are Presented

<table>
<thead>
<tr>
<th></th>
<th>ptk</th>
<th>b</th>
<th>d</th>
<th>g</th>
<th>glide</th>
<th>fric.</th>
<th>omis.</th>
<th>m/n</th>
<th>vowel</th>
</tr>
</thead>
<tbody>
<tr>
<td>p</td>
<td>11</td>
<td>11</td>
<td>34</td>
<td>34</td>
<td>24</td>
<td>17</td>
<td>22</td>
<td>22</td>
<td>1109</td>
</tr>
<tr>
<td>t</td>
<td>(.009)</td>
<td>7</td>
<td>105</td>
<td>50</td>
<td>(.030)</td>
<td>(.045)</td>
<td>18</td>
<td>15</td>
<td>(.009)</td>
</tr>
<tr>
<td>k</td>
<td>(.012)</td>
<td></td>
<td>(.030)</td>
<td></td>
<td>34</td>
<td>50</td>
<td>(.016)</td>
<td>(.014)</td>
<td>(.006)</td>
</tr>
<tr>
<td>b</td>
<td></td>
<td>13</td>
<td>(.030)</td>
<td></td>
<td>(.012)</td>
<td>43</td>
<td>10</td>
<td>26</td>
<td></td>
</tr>
<tr>
<td>d</td>
<td></td>
<td></td>
<td>(.015)</td>
<td></td>
<td>(.012)</td>
<td>14</td>
<td>(.013)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>g</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>21</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>glide</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(.019)</td>
<td>(.270)</td>
<td></td>
</tr>
<tr>
<td>fric.</td>
<td>24</td>
<td>17</td>
<td>96</td>
<td>123</td>
<td>21</td>
<td>62</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>omis.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(.111)</td>
<td>(.019)</td>
<td>(.056)</td>
</tr>
<tr>
<td>m/n</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>21</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>h</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(.019)</td>
<td>(.270)</td>
<td></td>
</tr>
<tr>
<td>r</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>28</td>
<td>(.025)</td>
<td></td>
</tr>
<tr>
<td>added</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>29</td>
<td>(.026)</td>
<td></td>
</tr>
<tr>
<td>vowel</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>31</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(.280)</td>
<td>(.022)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>433</td>
<td>(.080)</td>
<td></td>
</tr>
</tbody>
</table>
correlated with specific frequencies for Speaker 1. Instead, errors appeared to be equally distributed across frequencies.

Vowel substitutions were also analyzed (Table 12). These substitutions occurred 59 times and accounted for .8% of the total phoneme substitution errors.

Speaker 2

Table 10 illustrates that phoneme errors occurred more frequently for Speaker 2 (851) than for Speaker 1 (577).

Voicing substitution errors characteristic of Speaker 2 consisted of the plosives b/p (11%), d/t (9%), and g/k (7%). Substitution of voiced fricatives for voiceless fricatives comprised 15% of the total errors.

Nasal substitutions (9% of total errors) for the plosives /p/, /t/, and /k/ were also observed.

The most common phoneme error observed for Speaker 2 was the substitution of a voiced fricative for a voiceless fricative (15%). Other errors consistently noted were substitutions of the phonemes b/p and d/t (11%). Omission of phonemes /h/ (5%), /r/ (3%) were also commonly observed.

Vowel substitutions were more prevalent for Speaker 2 than for Speaker 1 (Table 12). A total of 225 (3%) vowels were substituted for other vowels.

Speaker 3

The phoneme errors most characteristic of Speaker 3 consisted
of substitutions of voiced phonemes for voiceless phonemes (Table 11, Appendix M). The most common error was the substitution of voiced fricatives for voiceless fricatives (11%). Substitution errors of the plosives d/t (10%), b/p (8%), and b/fricatives (9%) were also noted.

Errors made by the substitution of nasals of the phonemes /p/, /t/, /k/, /b/, /d/ and the fricatives /ʃ/, and /ð/ were also prevalent (16% of the total substitution errors).

Vowel substitutions were more characteristic of Speaker 3 than for Speakers 1 and 2. Vowel substitutions comprised 6.5% of the errors produced by Speaker 3 as compared to .7% for Speaker 1 and 3% for Speaker 2 (see Table 12).

**Summary**

Trends observed for Speakers 1 and 3 consisted of phoneme error substitutions of voiced for voiceless plosives and fricatives. Speaker 2 characteristically substituted the phonemes /p/, /t/, /b/ and /k/ for the fricatives. These comprised 22% of the total errors. Omission of the phoneme /h/ (5%) and substitution of the nasals /m/ and /n/ for the phonemes /p/, /t/, /k/, /b/, and /d/ occurred. This substitution pattern was observed a total of 172 times and accounted for 16% of the total errors.

Vowel substitutions were more common for Speaker 3 (6.5%) than for Speakers 2 (3%) and 1 (.8%).
Table 12

Percentage of Vowel Substitution Errors For Speakers Occuring at Each Frequency

<table>
<thead>
<tr>
<th></th>
<th>A(94 Hz.)</th>
<th>B(120 Hz.)</th>
<th>C(160 Hz.)</th>
<th>D(212 Hz.)</th>
<th>E(245 Hz.)</th>
<th>Total#</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speaker 1</td>
<td>35.6%</td>
<td>22.0%</td>
<td>18.6%</td>
<td>11.9%</td>
<td>11.9%</td>
<td>59</td>
</tr>
<tr>
<td>Speaker 2</td>
<td>11.5%</td>
<td>23.1%</td>
<td>21.3%</td>
<td>24.4%</td>
<td>19.5%</td>
<td>225</td>
</tr>
<tr>
<td>Speaker 3</td>
<td>15.7%</td>
<td>26.5%</td>
<td>14.5%</td>
<td>21.9%</td>
<td>21.2%</td>
<td>433</td>
</tr>
</tbody>
</table>
CHAPTER IV

DISCUSSION

Mean intelligibility scores of subjects listening to each speaker across five fundamental frequencies were determined. Intelligibility scores were examined for statistically significant differences between a) Speakers across frequencies and b) frequencies within Speakers for pooled and individual group means. The implications of these findings are discussed below. Other comparisons of error substitutions, speaker-acceptability, and sex recognition of speakers are also discussed in terms of the intelligibility differences observed.
The Relationship Between Fundamental Frequency ($F_0$) Of Electrolarynx Vibration And Speech Intelligibility Scores For Speakers

**Discussion of Group Order Effect**

The observed listening group vs. frequency interaction ($p=.026$) indicates that the responses of the 3 groups were not uniform for frequencies. This interaction would generally indicate that pooling of the groups for statistical comparisons might reduce the probability of observing frequency differences if they exist. However, the groups were pooled because consistent patterns for each speaker were observed across groups despite the statistically defined group-speaker interaction. These trends can be seen on Table 7.

It is difficult to ascertain the source of the interaction effect. However, it is possible that the characteristics of the listening groups themselves, the time at which they participated, or the order of presentation of stimulus items to each may have contributed to the interaction.

In regard to the latter possibility the presentation order of the frequencies was different for each speaker, but all listeners within a listening group was presented with the same order of frequencies for a given speaker. In other words, the order of presentation of frequencies was mixed for speakers but not randomized for individual listeners within listening groups.

Listener's acceptability ratings may also have been influenced by the order of presentation of the frequencies to each listener. Results indicated that the intelligibility ratings and listener's
acceptability seem to be related. That is, those frequencies yielding the most intelligible speech for each speaker were also rated as most acceptable. Table 13 indicates that presentation order within groups, rank of intelligibility scores, and acceptability of frequencies within speakers may be correlated to some degree. However, any consistent relationship between order and intelligibility is not immediately apparent. For example, Table 13 indicates that some relationship between order and rank of intelligibility scores for Speakers 2 and 3 exists. However, if such a relationship for Speaker 1 exists, the correlation would be negative.

The observed interaction between listening groups, speakers, and order of presentation of frequencies indicates that future designs should consider a) random presentations of the independent variables (speakers, frequencies) to each subject and b) participation within controlled time schedules.

**Between Speaker Intelligibility Comparisons**

Comparisons of intelligibility scores between speakers indicated that all three speakers were significantly different from each other at each of the five test frequencies. Speaker 1 was more intelligible than the other 2 speakers. One interpretation of the higher intelligibility of Speaker 1 might suggest that other within-speaker variables independent of frequency may have provided perceptual cues for listeners which enhance overall intelligibility for some speakers. Since the effects of speech rate, word stress and linguistic content should be minimized with this experimental design, it is
| Table 13 |
|------------------|------------------|------------------|
| **Rank Order of Intelligibility Scores and Acceptability Ratings (Low-High) and Order of Frequency Presentation for Each Group and Speaker. Frequency A=94Hz., Frequency B=120Hz., Frequency C=160Hz., Frequency D=212Hz., and Frequency E=245Hz.** |
| **Group 1** | **Speaker 1** | **Speaker 2** | **Speaker 3** |
| **Rank Order of Intelligibility Scores** | AEBCD | CDABE | BEDCA |
| **Acceptability** | EDCBA | CBDAE | EBDAC |
| **Order of Frequency Presentation** | ACBED | DEABC | EBDCA |
| **Group 2** | **Speaker 1** | **Speaker 2** | **Speaker 3** |
| **Rank Order of Intelligibility Scores** | ACEBD | DCBAE | BDEAC |
| **Acceptability** | EDCBA | DCBEA | BECDA |
| **Order of Frequency Presentation** | ACBED | DEABC | EBDCA |
| **Group 3** | **Speaker 1** | **Speaker 2** | **Speaker 3** |
| **Rank Order of Intelligibility Scores** | ACBED | CDBAE | BDEAC |
| **Acceptability** | EADCB | CDBAE | BDCEA |
| **Order of Frequency Presentation** | ACBED | DEABC | EBDCA |
probable that other factors (tissue density, viscosity, placement of the electrolarynx, articulatory proficiency of the speaker, etc.) may have accounted for at least part of the inter-speaker intelligibility differences noted.

As mentioned above, a variable which might account for general intelligibility differences between speakers may include the speaker's general articulatory proficiency. Different individuals articulate more precisely and accurately than others. Articulation patterns used during normal speech may be closely related to articulation patterns used during speech with an electrolarynx following a laryngectomy. Therefore, consideration of the individual's articulatory proficiency before the laryngectomy may prove to be beneficial in the planning and development of intervention strategies for post-surgical speech therapy.

Other variables which may account for inter-speaker differences of intelligibility may be related to the coupling of the artificial larynx to the user's neck. Coupling deficiency due to factors independent of frequency of vibration may include differences related to placement of the vibrating diaphragm on the user's neck or firmness of contact with the neck tissues which result in poor transmission of sound energy through the neck to the resonating vocal tract. Optimal placement for the laryngectomee is dependent upon tissue alterations (open fistulas, scarring, etc) which result from surgical intervention and may have very significant effects on the transfer characteristics of the filter system which effect the source-filter acoustic interactions,
physiological articulatory adjustments of the resonating system, and speech intelligibility.

Other speaker dependent coupling variables may also effect intelligibility. The noise produced by the electrolarynx which is not transmitted into the vocal tract may be enhanced when coupling is less efficient. While the ratio of the intensity of this "buzzing" noise to the speech signal may not be sufficient to impede the intelligibility of the signal itself, it may be of sufficient intensity to promote psychophysiological interference which alters variables critical for speech perception and speech intelligibility (Barney, Haworth, Dunn, 1959). In that regard, it is interesting to consider the finding that acceptability and intelligibility rating seem to be highly related. This may indicate that a) the frequency cues used for speech perception may also be related to cues determining acceptability of alaryngeal speech or b) the reduced intelligibility in general decreases the listener's perceptual "set" for attending to the speaker's utterances. If the listener has difficulty understanding a speaker this may have some effect on his attitude toward or acceptance of that speaker. If the speaker is not acceptable to a listener, intelligibility may be effected even though acoustic speech signals are adequate. It is difficult to determine whether the variables correlated with intelligibility and acceptability covary, or whether variables related to one have some systematic effect on the other. These possibilities must be subjected to empirical test.
Intelligibility Scores For Frequencies Within Speakers

Intelligibility scores within speakers varied across frequencies. Maximum intelligibility scores were more highly correlated with certain frequencies for certain speakers. Speaker 1 (female, 5'3" tall, 100 lbs., fundamental frequency of 190 Hz.) obtained the highest intelligibility scores when the electrolarynx was set at 212 Hz. and 120 Hz. and the lowest intelligibility scores at 94 Hz. Speaker 2 (female, 5'8" tall, 135 lbs., fundamental frequency 230 Hz.) obtained significantly higher intelligibility scores at 245 Hz. and significantly lower scores at 160 Hz. and 212 Hz. Speaker 3 (male, 6'4", 190 lbs., fundamental frequency 100 Hz.) obtained significantly higher intelligibility scores at 160 Hz. and 94 Hz. and significantly lower scores at 120 Hz.

While vocal tract dimensions were not measured for speakers, it was assumed that general physical differences between speakers also reflected inter-speaker differences of vocal tract dimensions. The Speaker's physical structure has some effect on the interaction of the artificial vibrating source and the vocal resonant system. Theoretically, the size of the resonating cavity, the amount of constriction at the opening of the vocal tube, and the length of constriction all interact to determine the precise natural resonant frequencies of the tube (Stevens and House, 1955). Therefore, the volume and tissue characteristics of the resonating cavity are closely related to the resonances which will be produced (Minifie, 1973). The match of resonant frequencies of the tract to the frequency of
the vibrating source will be related to the efficacy of the speech
sound process.

While the dynamic changes of the vocal tract, volume, degree of
constriction or location of the constriction is very complex during
speech, the system will transmit a speech signal most efficiently if
the source and resonant characteristics are compatible. It has been
demonstrated that the three-dimensional shape of the vocal tract
at any instant in time is the primary determinant of the signal
being produced. The vocal tract serves to selectively modify the
acoustical signal from the sound source so that the sound at the
mouth opening is dependent upon both the nature of the sound source
and the nature of the acoustical filter (Minifie, 1973). In this
investigation, the speaker's acoustical filter (vocal tract) interacted
differently when the parameter of frequency of vibration of the source
was altered. Several frequencies appeared to enhance the efficiency
(intelligibility) of the resonating system by increasing intelligibility,
while other frequencies appeared to promote less satisfactory resonances
for speech intelligibility. The exact relationship between the absolute
values of most or least intelligible frequencies and the speaker's
natural fundamental frequency or vocal tract characteristics is not
immediately apparent in this investigation. However, it is probable
that source-filter compatibility altered the tuning of the system
in some way. The number or values of the frequencies chosen for
this experiment may not have been appropriate for the determination
of optimum harmonic/filter matchings for each speaker. More frequencies
should be evaluated for individual speakers to determine the
relationship between the Speaker's fundamental frequency, vocal tract resonances of the Speaker, and intelligibility. It was demonstrated however, that intelligibility scores did vary as a function of frequency for all 3 speakers. This indicates that one, two or more frequencies may yield more intelligible speech for a speaker, and these "intelligible" frequencies differ for different speakers. This finding should be taken into consideration when the selection of an appropriate artificial larynx is considered for the laryngectomized individual.

Analysis Of Phoneme Errors By Frequency

Statistical analysis of specific error types was not conducted for this experiment. However, descriptive data are provided for discussion. Errors of manner (plosive, fricative, affricate, liquid, glide, nasal) and place (bilabial, labiodental, lingualvelar and velar) were calculated and matrices were developed to determine the percentage of phoneme error types and error trends for each speaker at each frequency.

For all three speakers, fricatives were the most poorly perceived, while stops and affricates were the next most poorly perceived classes. These results are in agreement with those obtained by Weiss and Basili (1981) in an investigation analyzing phoneme errors characteristic of electrolarynx speech. Discussion of the relation of phoneme error substitution types to speech intelligibility and to frequency of vibration is provided below.
Errors of Voicing

The substitution of voiced for voiceless phonemes was the most prominent phoneme error substitution observed for all speakers and across all frequencies. Specifically, phoneme substitutions of b/p, d/t, g/k and voiced for voiceless fricatives comprised the majority of voicing error substitutions. As demonstrated by Kuhn (1975), primary cues for voicing for stops are the presence or absence of concentrations of sound energy at low frequencies, the presence or absence of noise indicating aspiration, or a change in voice onset time. It has been concluded that listeners seem to make some decisions about voicing on the basis of timing patterns, especially those patterns of voice onset time (Lisker and Abramson, 1965). The findings of the present investigation indicating frequent phoneme error substitutions of stop plosives may indicate that timing patterns may have been altered as test frequencies varied, or the temporal relation of plosive noise bursts to the onset of sound may have changed at the different test frequencies. It is interesting to note that while certain frequencies were more intelligible for certain speakers, significantly more voicing errors for stops were observed at 94 Hz. for all 3 speakers. This finding suggests that subsequent investigations of the relationship between temporal acoustic factors and specific frequencies of vibration must be completed to determine critical variables underpinning such a relationship. If the acoustic cues for consonant voicing depend more upon relative durations and timing of events than upon frequency or intensity differences as suggested by
Denes (1955), and Liberman, Delattre, and Cooper (1958), one would not expect to see such a large number of voicing errors across speakers which are common to one frequency of the vibrating source. Therefore, the contribution of other cues such as the relationship between individual formant frequencies, the relative intensity of formants, or bandwidths of these formants must be determined. Additionally, alterations of the frequency of vibration of a sound source may alter the timing and duration cues used by the listener to differentiate between voiced and voiceless stops. Phoneme or syllabic durations were not measured in this investigation, but their effects should also be studied in follow-up analyses.

In addition to the high frequency of occurrence of substitution errors of voiced for voiceless stops, the substitution of voiced for voiceless fricatives and affricates was also commonly observed for all speakers. No pattern or relationship between these error types and the most intelligible frequency of vibration were observed. According to Denes (1955) the detection of voicing for fricatives depends upon the presence of energy concentrations at the low frequencies of the sound spectra. This remains a salient cue, but even without it, listeners can make judgments about the voicing of a syllable-final fricative based upon its duration relative to the duration of the preceding vowel. Electrolarynx speech is characterized by a constant low frequency concentration of energy. This continuous acoustic signal accompanying both voiced and voiceless fricative productions may interfere with listeners' perception of duration cues used to distinguish between different fricatives.
By altering the frequency of vibration of the sound source, changes in articulatory patterns, and thus vowel durations may also be observed. These changes in durations may influence the perception of a fricative which follows a vowel. Harris (1958) observed that listeners relied on the changing formant structures stemming from changes of duration in adjacent vowels to determine whether they heard /i/ or /e/. Consideration of Harris' findings in light of the findings of the present investigation suggests further investigation of the relation of vowel duration to frequency, intensity, etc. to determine the effects of those variables on listeners' perception of speech produced with an electrolarynx or produced normally.

Substitutions of Vowels

Vowel substitutions were characteristic of errors across all five frequencies for all 3 speakers. While not analyzed statistically, calculation of the absolute number of vowel substitutions across frequencies indicate that vowel substitutions seemed to occur less frequently at more intelligible frequencies. Alterations of the frequency of the vibrating source may have altered the acoustic cues used in the perception of vowels to be altered. This speculation is supported by the work of Delattre, Liberman and Cooper (1955) which indicated that the cues to the perception of vowels lie in the patterns created by the vocal tract resonances (formants) of the speakers. These formants reflect energy concentrations at different frequencies and may be altered with variations of the frequency of the vibrating source and therefore may interfere with those formant patterns normally
created by the vocal tract at its natural resonant frequency.

Delattre, Liberman, and Cooper (1966) also concluded that listeners use general patterns for formant relationships rather than absolute frequency values or even an exact ratio of frequencies. We might expect to see changes in formant characteristics when the frequency of the voicing source is altered. Less dramatic changes of intelligibility may be observed when the frequency of electrolarynx vibration more closely approximates the system's natural resonating frequencies.

Although vowel perception errors related to vowel production are rather infrequent in normal speech production, they seem to play a more prominent role in the intelligibility of electrolarynx speech. The high occurrence of vowel substitutions in this experiment indicates that the parameter of frequency may be critical for enhancing the perception of vowels in electrolarynx speech, at least for some individuals.

Errors in Manner of Production

Manner contrasts (fricative, stop, nasal etc.) appear to be based upon the parameters of frequency, intensity and timing. The alteration of frequency in this study may have had an effect on the number and type of errors of manner of production observed for 2 of the 3 speakers. Since errors of manner characterized the scores for two speakers but not the 3rd, the relationship of frequency change and errors of manner might be considered to be a speaker dependent variable. In order to further discuss the relationship
between fundamental frequency and errors of manner, the perceptual cues used by a listener to determine manner of production of a phoneme are described.

Borden and Harris (1980) suggest that when listeners identify the manner of production of a speech sound, they determine whether the sound is harmonically structured without speech noise (which signals the production of vowels, semivowels, or nasals) or whether the sound contains an aperiodic noise component (which signals the production of stops, fricatives or affricates). Additionally, the periodic, harmonically structured classes present acoustic cues in energy regions that are relatively low in frequency, while the aperiodic, noisy classes of speech sounds are cued by energy that is relatively high in frequency. If the low frequency components are not enhanced because of poor source-filter matchup, they may be more likely to be distorted or masked by other signals. This may lead to a sufficient disruption of the speech signal and may promote the listener's misperceptions of a stop or fricative.

In addition, the relative intensity of individual formants and absolute formant frequency changes are primary manner cues available for a listener to distinguish between vowels, nasals and semivowels. The nasal consonants have formants which are less intense than formants for semivowels and vowels, and they possess a distinctive low frequency resonance (Malecot, 1956). This low frequency resonance characteristic of nasal consonants may be confused with the concentration of energy at the low frequencies which is characteristic of the electrolarynx vibration system. Thus, the perceptual cue that listeners'
use to perceive nasal consonants may be altered significantly in electrolarynx speech.

Duration also may have contributed to manner substitution errors of stops, fricatives and affricates in this study. The duration of the noise, which is transient for stops, longer for affricates, and longest for fricatives is a manner cue associated with sounds having an aperiodic component. These duration cues may have been altered as a result of changing the frequency setting of the electrolarynx. However, alteration of the frequency of the vibrating source did not seem to interfere with the perception of manner for all 3 speakers. Apparently, the recordings of Speaker 1 for some reason did not yield phoneme error substitutions of manner to any notable degree (see Table 9). This may indicate that errors resulting in manner of production substitutions may be determined by unique individual speaker-frequency interactions. For example, more substitution errors of nasals for stops and fricatives were observed at 212 Hz. for Speakers 2 and 3 while 120 Hz., 245 Hz., and 160 Hz. yielded more substitution errors of stops for fricatives for those speakers. These results indicate that errors of manner may be speaker or frequency dependent. The analysis of such relationships may eventually be an important criterion for determination of an appropriate device.

Errors in Place of Articulation

Alterations of the frequency of the electrolarynx may have contributed to error substitutions of place such as the alveolar
for velar (t/k) or alveolar for labial (d/p) substitutions observed for Speakers 2 and 3 (Tables 10-11). According to Delattre, Liberman, and Cooper (1955) the acoustic cues for place of articulation depend more upon the parameter of frequency than on duration or intensity. More specifically, two acoustic cues to place of articulation for the location of stops, fricatives and affricates are the second formant ($f_2$) transitions into neighboring vowels and the frequency of the noise components. Consequently, a transition of a second formant with a low locus (frequency) is perceived as labial. With a higher locus, it is alveolar, and with a varied, vowel-dependent locus, it is perceived as a palatal or velar.

Alteration of the frequency of the vibrating source may have introduced changes or disturbances in the formant transitions themselves and the relative position of the locus of the second formants of the speech signal. These disruptions in transitional cues may have been one possible determinant of the substitutions of stops for fricatives observed for all 3 speakers in this study. This possibility should be investigated.

Substitution errors of fricatives for other fricatives ($\theta/f$, $\eta/v$, $\theta/v$, $\eta/f$) comprised the majority of errors of place of articulation. This is not surprising based on the literature investigating the acoustic cues present for fricative discrimination. According to Harris (1958), the frequency of the noise itself is a cue for place of production. Raphael and Dorman (1977) report that if the friction covers a wide band of frequencies, it is more likely to be perceived as /f/, /θ/ or /h/.
The low intensity friction noise which characterizes the /θ/, 
/ð/, /f/ and /v/ phonemes may have been distorted or masked by the 
low frequency noise produced by the electrolarynx in some way. This 
speculation is supported by results of the investigation of Miller and 
Nicely (1955) which demonstrated that /v/ and /ð/ were among the 
most confusable of speech sounds to listeners when noise was added 
to the stimuli.

Sex Recognition of Speakers Using An Artificial Larynx

Results of this investigation indicated that listeners may 
use frequency related cues to determine the sex of a speaker. 
Listeners' consistently perceived the two lowest speaking frequencies 
(94 Hz. and 120 Hz.) as being produced by male speakers, and the 
highest two frequencies (212 Hz., and 245 Hz.) as being produced 
by female speakers (Table 14). This is not surprising since the 
average fundamental frequency for males is approximately 120 Hz. 
and for females approximately 212 Hz. This sex recognition-frequency 
relationship should be considered when selecting a frequency setting 
for the laryngectomee. If there are several frequencies which appear 
to yield approximately the same intelligibility scores, consideration 
should be given to the frequency which most appropriately reflects 
the speaker's sex, especially since acceptability seems to be related 
to intelligibility.

Acceptability Ratings of Speakers

Information about the ultimate acceptability or various types of
Table 14
Percentage of Listeners in Each Group That Indicated the Speaker was Male

<table>
<thead>
<tr>
<th>Frequency</th>
<th>A(94Hz.)</th>
<th>B(120Hz.)</th>
<th>C(160Hz.)</th>
<th>D(212Hz.)</th>
<th>E(245Hz.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speaker 1</td>
<td>28.6%</td>
<td>42.8%</td>
<td>14.3%</td>
<td>14.3%</td>
<td>28.6%</td>
</tr>
<tr>
<td>Group 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Speaker 2</td>
<td>100%</td>
<td>85.5%</td>
<td>14.3%</td>
<td>0.0%</td>
<td>14.3%</td>
</tr>
<tr>
<td>Speaker 3</td>
<td>100%</td>
<td>100%</td>
<td>42.8%</td>
<td>85.8%</td>
<td>57.1%</td>
</tr>
<tr>
<td>Speaker 1</td>
<td>85.8%</td>
<td>42.8%</td>
<td>0.0%</td>
<td>28.5%</td>
<td>0.0%</td>
</tr>
<tr>
<td>Group 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Speaker 2</td>
<td>85.8%</td>
<td>42.8%</td>
<td>14.3%</td>
<td>14.3%</td>
<td>14.3%</td>
</tr>
<tr>
<td>Speaker 3</td>
<td>100%</td>
<td>100%</td>
<td>85.8%</td>
<td>57.1%</td>
<td>71.5%</td>
</tr>
<tr>
<td>Speaker 1</td>
<td>100%</td>
<td>100%</td>
<td>57.1%</td>
<td>28.6%</td>
<td>28.6%</td>
</tr>
<tr>
<td>Group 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Speaker 2</td>
<td>71.5%</td>
<td>75.5%</td>
<td>57.1%</td>
<td>42.8%</td>
<td>0.0%</td>
</tr>
<tr>
<td>Speaker 3</td>
<td>100%</td>
<td>85.8%</td>
<td>57.1%</td>
<td>42.8%</td>
<td>57.1%</td>
</tr>
</tbody>
</table>

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
alaryngeal speech is important for the laryngectomized patient and for his listener. Experiments have been performed to specify vocal attributes which differentiate good and poor esophageal speakers (Snidecor and Curry, 1959; Shipp, 1967; Hoops and Noll, 1969), and which allow comparisons of listener preferences of esophageal and artificial larynx speech (Hyman, 1955; Crouse, 1962; Snidecor, 1968). However, there has been little or no research specifically investigating acoustic variables which alter the acceptability of electrolaryngeal speech. A recent experiment investigating and comparing the acceptability ratings of normal, esophageal, and artificial larynx speech was conducted by Bennett and Weinberg (1973). In that study, judges rated the acceptability of a number of subject's speech using a seven-point equal-appearing interval scale, where 1 represented speech which was least acceptable and 7 represented speech which was highly acceptable. The nature of the voices to be rated was not presented. Listeners were given the following specific information concerning speech acceptability (Bennett and Weinberg, 1973):

In making your judgments about the speakers you are about to hear, give careful consideration to the attributes of pitch, rate, understandability, and voice quality. In other words, is the voice pleasing to listen to, or does it cause you some discomfort as a listener? (pg. 453).

Results of this study indicated that speech with an artificial larynx was less acceptable than speech using esophageal methods. However, the Bennett and Weinberg study failed to recognize that different speakers implementing the same form of alaryngeal speech may receive different acceptability ratings from different listeners. In the present study, instructions to the listeners were identical to
those used in the aforementioned study (Bennett and Weinberg, 1973).

Results suggest a relationship between the mean percentage intelligibility scores and the mean acceptability rating scores for each listener (The 1-7 rating scale was converted to percentages on the basis of a 100 point scale) for frequencies across speakers for all three groups.

Speaker 1 was rated more acceptable than Speaker 2 and 3 by 2 of the 3 listening groups. Table 13 suggests that a relationship between the frequencies yielding the highest intelligibility scores and the frequencies yielding the highest acceptability rating exists for each Speaker. Similar relationships were observed between the frequencies yielding the lowest intelligibility scores and the frequencies yielding the lowest acceptability rating scores (see Table 13).

These preliminary descriptive data are important when considering additional research related to factors which may play a role in speech intelligibility. Listener acceptability judgments are in some way related to perceived intelligibility of a speaker. It is obvious that speech acceptability changes when the fundamental frequency of the artificial sound source is altered. No single frequency of electrolarynx vibration was observed to yield higher acceptability rating across individual speakers or across listening groups.

The apparent correlation between intelligibility and acceptability suggests that changes of intelligibility may effect the psychodynamic (acceptance/rejection) relationships between speaker and listener and may magnify the communicative disruption related to reduced intelligibility. Certainly, the cosmetic effect and qualitative
listener judgments of electrolarynx speech play an important role in the speaker's motivation to use such devices or in the listener's acceptance of the alaryngeal speaker. Therefore, acceptability as well as intelligibility should be taken into consideration for each individual when selection of an electrolarynx or an appropriate frequency of electrolarynx vibration is being conducted. Acceptability and intelligibility factors should be evaluated on a within subject basis.
CHAPTER V
SUMMARY AND CONCLUSIONS

Purpose

This study was designed to test the hypothesis that speech intelligibility varies systematically across speakers and fundamental frequency of electrolarynx vibration. Mean intelligibility scores were obtained on the three groups of listeners for three speakers across five frequencies (94 Hz., 120 Hz., 160 Hz., 212 Hz., and 245 Hz).

Comparative analysis of scores across frequencies and speakers were discussed in terms of:

a) source transmission characteristics of "electrolaryngeal" speech,

b) implications for diagnosis/therapeuis of the alaryngeal speakers,

c) acceptability of alaryngeal speech,

d) sex recognition of alaryngeal speakers, and

e) phoneme error substitution types and frequency of electrolarynx vibration.

Experimental Design

Speakers

Two female adults and one male adult were selected as speakers for this study. Speakers had normal hearing, normal articulation, and had no previous training with the use of an artificial larynx. Speakers also represented a continuum of physiological and anatomical differences based upon sex and physique. Assuming that individuals

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
of different stature would differ with regard to vocal tract size and sound transmission characteristics this selection criterion was used to obviate intelligibility differences which might be related to the coupling of various fundamental frequencies to various size vocal tracts if such differences exist. The ages of subjects ranged from 22 years to 26 years (\( \bar{X} = 24.5 \) years).

Stimuli

Five lists of PB-50 words embedded in a carrier phrase were recorded by each speaker using a Western Electric No. 5 electric larynx.

Procedures

Recording of Stimulus Words

Three speakers made recordings of 5 CVC word lists using a Western Electric No. 5 electric larynx. A brief training period for the speakers was conducted to instruct them with the operation of the electrolarynx. Following the training segment, each speaker was instructed to produce each PB word by attaching it to the carrier phrase "You will say ______." Each speaker produced a 50 word list for each test frequency.

Presentation of Stimulus Words

Three listening groups (N=7) participated in the listening task. Each listening group was presented with five sets of 50

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
CVC word lists for each of three speakers at each of five frequencies. Presentation of speakers and frequencies were ordered differently for each listening group.

When the intelligibility portion of the experiment was completed, listeners were instructed to make judgments of acceptability and sex recognition of Speaker.

Findings and Conclusions

Findings

Findings of the study provided the following answers to the experimental questions:

1. Speech intelligibility scores were significantly different for different frequencies of electrolarynx vibration within speakers.

2. Intelligibility scores for individual frequencies were significantly different across speakers.

Conclusions

Results of this study indicate that both mechanical and acoustic transmission characteristics of the artificial larynx must be evaluated more completely to maximize the efficiency of these systems for each individual. Further studies investigating the relationship of intelligibility of electrolaryngeal speech and fundamental frequency, duration, or intensity characteristics of electrolarynx speech should be conducted.

Since this investigation indicated that fundamental frequency of
electrolarynx vibration and intelligibility are related, and since this relationship varies across speakers, it seems appropriate to expend additional research efforts to determine the effects of:

a) the relationship between acceptability and intelligibility,

b) characteristics of phoneme error substitutions most commonly observed for different speakers using an electrolarynx,

c) the relationship of formant structure to fundamental frequency of vibration and to speech intelligibility for electrolarynx speakers,

d) changes of duration resulting from alterations of frequency of the vibrating sound source,

e) changes of intensity with fundamental frequency alterations,

f) the relationship of frequency and durations which may play a role in sex recognition of speakers using an electrolarynx,

g) the energy transmission characteristics resulting from the coupling of the vibrating source with tissues in the neck region,

h) the variations of articulatory dynamics across fundamental frequencies of vibration for speakers using an electrolarynx.

Because there was some interaction between listening groups, speakers, and the order of presentation of frequencies to listeners, future research designs should consider the effects of practice, perceptual set of the listener etc., on the independent variables of fundamental frequency, vocal tract size, etc.

Results of this investigation indicate that alterations in one variable (frequency) may result in differences in intelligibility, acceptability, phoneme error substitutions, and sex recognition of the speaker. Immediate clinical implications may include:

a) Circuitry alterations which promote variable fundamental frequency settings to determine frequencies which would yield maximum intelligibility for a speaker,
b) Alteration of the frequency of the electrolarynx to a more acceptable pitch level according to the sex of the speaker.

In conclusion, the goal of the speech clinician working with the laryngectomized individual should be to acquire the most intelligible, and satisfactory system which is most acceptable to those in the speaker's environment. It seems that we have been too accepting of the electronic systems which are presently available for the laryngectomized individual. Research energy must be channeled to this communicatively impaired population and to the alternative systems which will best enhance their ability to communicate effectively. The present investigation suggests that we can significantly improve the intelligibility and the acceptability of speech produced with an electrolarynx by investigating those variables which will promote modifications in the prosthetic device to maximize intelligibility of the laryngectomized speaker.
APPENDICES
Appendix A

PB-50 Word Lists
Selected From The USA Monosyllabic Intelligibility Test Word Lists

<table>
<thead>
<tr>
<th>List 1</th>
<th>List 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 came</td>
<td>23 death</td>
</tr>
<tr>
<td>2 there</td>
<td>24 are</td>
</tr>
<tr>
<td>3 dish</td>
<td>25 bad</td>
</tr>
<tr>
<td>4 hid</td>
<td>26 pest</td>
</tr>
<tr>
<td>5 heap</td>
<td>27 slip</td>
</tr>
<tr>
<td>6 pants</td>
<td>28 rub</td>
</tr>
<tr>
<td>7 hunt</td>
<td>29 feast</td>
</tr>
<tr>
<td>8 no</td>
<td>30 deed</td>
</tr>
<tr>
<td>9 bar</td>
<td>31 cleanse</td>
</tr>
<tr>
<td>10 pan</td>
<td>32 fold</td>
</tr>
<tr>
<td>11 fuss</td>
<td>33 nook</td>
</tr>
<tr>
<td>12 creed</td>
<td>34 mange</td>
</tr>
<tr>
<td>13 box</td>
<td>35 such</td>
</tr>
<tr>
<td>14 strife</td>
<td>36 use (yews)</td>
</tr>
<tr>
<td>15 dike</td>
<td>37 crash</td>
</tr>
<tr>
<td>16 not</td>
<td>38 ride</td>
</tr>
<tr>
<td>17 ford</td>
<td>39 pile</td>
</tr>
<tr>
<td>18 end</td>
<td>40 rat</td>
</tr>
<tr>
<td>19 then</td>
<td>41 rag</td>
</tr>
<tr>
<td>20 bask</td>
<td>42 is</td>
</tr>
<tr>
<td>21 fraud</td>
<td>43 wheat</td>
</tr>
<tr>
<td>22 smile</td>
<td>44 rise</td>
</tr>
</tbody>
</table>

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
<table>
<thead>
<tr>
<th>List 3</th>
<th>List 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 why</td>
<td>26 size</td>
</tr>
<tr>
<td>2 turf</td>
<td>27 wedge</td>
</tr>
<tr>
<td>3 gnaw</td>
<td>28 deck</td>
</tr>
<tr>
<td>4 drop</td>
<td>29 hurl</td>
</tr>
<tr>
<td>5 jam</td>
<td>30 wharf</td>
</tr>
<tr>
<td>6 flush</td>
<td>31 leave</td>
</tr>
<tr>
<td>7 rouse</td>
<td>32 crave</td>
</tr>
<tr>
<td>8 neck</td>
<td>33 vow</td>
</tr>
<tr>
<td>9 sob</td>
<td>34 law</td>
</tr>
<tr>
<td>10 trip</td>
<td>35 stag</td>
</tr>
<tr>
<td>11 dill</td>
<td>36 oak</td>
</tr>
<tr>
<td>12 thrash</td>
<td>37 nest</td>
</tr>
<tr>
<td>13 dig</td>
<td>38 sit</td>
</tr>
<tr>
<td>14 rate</td>
<td>39 crime</td>
</tr>
<tr>
<td>15 far</td>
<td>40 muck</td>
</tr>
<tr>
<td>16 check</td>
<td>41 fame</td>
</tr>
<tr>
<td>17 air</td>
<td>42 take</td>
</tr>
<tr>
<td>18 bead</td>
<td>43 who</td>
</tr>
<tr>
<td>19 sped</td>
<td>44 toil</td>
</tr>
<tr>
<td>20 cast</td>
<td>45 path</td>
</tr>
<tr>
<td>21 class</td>
<td>46 pulse</td>
</tr>
<tr>
<td>22 lush</td>
<td>47 fig</td>
</tr>
<tr>
<td>23 shout</td>
<td>48 barb</td>
</tr>
<tr>
<td>24 bald</td>
<td>49 please</td>
</tr>
<tr>
<td>25 cape</td>
<td>50 ache</td>
</tr>
</tbody>
</table>

Appendix A—Continued
Appendix A—Continued

List 5

1 feed 25 shine 49 nose
2 gape 26 sly 50 grudge
3 sick 27 wrath
4 Greek 28 love
5 roe 29 beck
6 choose 30 thick
7 true 31 flap
8 pass 32 cheat
9 browse 33 wink
10 punt 34 zone
11 shove 35 odds
12 hill 36 kid
13 black 37 trade
14 high 38 scare
15 rind 39 mast
16 vase 40 pipe
17 rode 41 good
18 puff 42 lend
19 inch 43 yawn
20 bronze 44 watch
21 solve 45 thud
22 bathe 46 tug
23 add 47 curse
24 rear 48 owls

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Appendix B

Schematic of Adjusted Electrolarynx Modified Switch Circuitry

Heath L.V.
Power Supply
Model IP-27

(10V, 150ma, setting)

To Armature of Pitch Control
To Neutral Terminal
To Battery Terminal

To Battery Terminal
To Neutral Terminal
To Armature of Pitch Control
### Appendix C

**Sound Pressure Levels for Unloaded and Loaded Systems at Five Test Frequencies**

<table>
<thead>
<tr>
<th>Frequency Setting</th>
<th>Intensity For Unloaded Condition</th>
<th>Intensity For Loaded Closed Condition</th>
<th>Intensity For Loaded Open Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>A 94Hz.</td>
<td>85dB</td>
<td>62dB</td>
<td>74dB</td>
</tr>
<tr>
<td>B 120Hz.</td>
<td>85dB</td>
<td>63dB</td>
<td>74dB</td>
</tr>
<tr>
<td>C 160Hz.</td>
<td>87dB</td>
<td>65dB</td>
<td>73dB</td>
</tr>
<tr>
<td>D 212Hz.</td>
<td>88dB</td>
<td>66dB</td>
<td>77dB</td>
</tr>
<tr>
<td>E 245Hz.</td>
<td>89dB</td>
<td>68dB</td>
<td>72dB</td>
</tr>
</tbody>
</table>
Appendix D

Recording Order for Word Lists (1, 2, 3, 4, 5) and Frequencies 94 Hz. (A), 120Hz. (B), 160Hz. (C), 212Hz. (D), 245Hz. (E) for Speaker 1, Speaker 2, and Speaker 3.

<table>
<thead>
<tr>
<th>Frequency Order</th>
<th>Speaker 1</th>
<th>Speaker 2</th>
<th>Speaker 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACBDE</td>
<td>DEABC</td>
<td>EBDCA</td>
<td></td>
</tr>
<tr>
<td>List Order</td>
<td>45123</td>
<td>52431</td>
<td>13254</td>
</tr>
</tbody>
</table>
### Appendix E

Rank Order of Intelligibility Scores and Acceptability Ratings (Low-High) and Order of Frequency Presentation for Each Group and Speaker. Frequency A=94Hz., Frequency B=120Hz., Frequency C=160Hz., Frequency D=212Hz., and Frequency E=245Hz.

<table>
<thead>
<tr>
<th>Group</th>
<th>Speaker 1 Rank Order of Intelligibility Scores</th>
<th>Speaker 2 Acceptability</th>
<th>Speaker 3 Order of Frequency Presentation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Rank Order of Intelligibility Scores</td>
<td>Acceptability</td>
<td>Order of Frequency Presentation</td>
</tr>
<tr>
<td></td>
<td>AEBCD</td>
<td>EDCBA</td>
<td>ACBED</td>
</tr>
<tr>
<td>Group 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ACBED</td>
<td>DCBAE</td>
<td>DEABC</td>
</tr>
<tr>
<td></td>
<td>ACBED</td>
<td>DCBAE</td>
<td>EBDECA</td>
</tr>
<tr>
<td>Group 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ACEBD</td>
<td>EDCBA</td>
<td>ACBED</td>
</tr>
<tr>
<td></td>
<td>ACBED</td>
<td>DCBDEA</td>
<td>DEABC</td>
</tr>
<tr>
<td></td>
<td>ACBED</td>
<td>DCBDEA</td>
<td>EBDEAC</td>
</tr>
<tr>
<td>Group 3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ACBED</td>
<td>EADCB</td>
<td>ACBED</td>
</tr>
<tr>
<td></td>
<td>ACBED</td>
<td>CDBAE</td>
<td>DEABC</td>
</tr>
<tr>
<td></td>
<td>ACBED</td>
<td>CDBAE</td>
<td>EBDEAC</td>
</tr>
</tbody>
</table>
Appendix F

Sound Spectra of the Vowel "ee" (as in "bean") Produced With the Electrolarynx
Set at the Five Test Frequencies
Appendix G

Sound Spectra of the Vowel Sound "a" (as in "father") Produced with the Electrolarynx Set At the Five Test Frequencies
Appendix H

Sound Spectra of the Electrolarynx Set at the Five Test Frequencies
94Hz., 120Hz., 160Hz., 212Hz., 245Hz.

<table>
<thead>
<tr>
<th>Frequency</th>
<th>94Hz</th>
<th>120Hz</th>
<th>160Hz</th>
<th>212Hz</th>
<th>245Hz</th>
</tr>
</thead>
<tbody>
<tr>
<td>8k</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7k</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6k</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5k</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4k</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3k</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2k</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1k</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Appendix I

**Rank Order of Mean Intelligibility Scores (Parenthesis) From Low-High and Their Corresponding Frequencies for Each Speaker Across Groups (Frequency A=94Hz., B=120Hz., C=160Hz., D=212Hz., and E=245Hz.)**

<table>
<thead>
<tr>
<th>Speaker 1</th>
<th>Speaker 2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Group 1</strong></td>
<td></td>
</tr>
<tr>
<td>A(70.0)</td>
<td>G(39.7)</td>
</tr>
<tr>
<td>E(73.4)</td>
<td>D(41.7)</td>
</tr>
<tr>
<td>B(74.5)</td>
<td>A(50)</td>
</tr>
<tr>
<td>C(77.4)</td>
<td>B(50.2)</td>
</tr>
<tr>
<td>D(79.4)</td>
<td>E(56.9)</td>
</tr>
<tr>
<td><strong>Group 2</strong></td>
<td></td>
</tr>
<tr>
<td>A(64.5)</td>
<td>D(34.6)</td>
</tr>
<tr>
<td>C(65.7)</td>
<td>C(36.3)</td>
</tr>
<tr>
<td>E(60.1)</td>
<td>B(50.9)</td>
</tr>
<tr>
<td>D(72.6)</td>
<td>A(54.6)</td>
</tr>
<tr>
<td>B(73.7)</td>
<td>E(55.1)</td>
</tr>
<tr>
<td><strong>Group 3</strong></td>
<td></td>
</tr>
<tr>
<td>A(50.9)</td>
<td>C(39.1)</td>
</tr>
<tr>
<td>C(59.4)</td>
<td>D(46.9)</td>
</tr>
<tr>
<td>B(62.3)</td>
<td>B(47.4)</td>
</tr>
<tr>
<td>E(64.0)</td>
<td>E(65.1)</td>
</tr>
<tr>
<td>D(65.1)</td>
<td>A(51.7)</td>
</tr>
<tr>
<td><strong>Pooled</strong></td>
<td></td>
</tr>
<tr>
<td>A(62.5)</td>
<td>C(38.4)</td>
</tr>
<tr>
<td>C(67.5)</td>
<td>D(41.0)</td>
</tr>
<tr>
<td>E(68.9)</td>
<td>B(49.5)</td>
</tr>
<tr>
<td>B(70.2)</td>
<td>A(52.1)</td>
</tr>
<tr>
<td>D(72.4)</td>
<td>E(58.4)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Speaker 3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Group 1</strong></td>
</tr>
<tr>
<td>B(27.4)</td>
</tr>
<tr>
<td>D(35.1)</td>
</tr>
<tr>
<td>C(42.3)</td>
</tr>
<tr>
<td><strong>Group 2</strong></td>
</tr>
<tr>
<td>B(32.3)</td>
</tr>
<tr>
<td>D(36.9)</td>
</tr>
<tr>
<td>A(42.6)</td>
</tr>
<tr>
<td>C(50.3)</td>
</tr>
<tr>
<td><strong>Group 3</strong></td>
</tr>
<tr>
<td>B(24.6)</td>
</tr>
<tr>
<td>D(31.7)</td>
</tr>
<tr>
<td>A(39.1)</td>
</tr>
<tr>
<td>C(42.0)</td>
</tr>
<tr>
<td><strong>Pooled</strong></td>
</tr>
<tr>
<td>B(28.1)</td>
</tr>
<tr>
<td>D(34.6)</td>
</tr>
<tr>
<td>A(42.7)</td>
</tr>
<tr>
<td>C(44.9)</td>
</tr>
</tbody>
</table>
Appendix J

Instructions for Speakers and Listeners

Listeners

You are about to participate in an experiment related to the understanding of speech signals produced by an artificial larynx.

You will be listening to many sentence productions which will be presented through a loudspeaker at a comfortable and safe level.

You will be listening to these sentences for approximately 1½ hours. During this time you will be given three, five-minute breaks to relax and stretch.

Each sentence will be initiated with the carrier phrase "You will say ______". You are to write down the last word of the sentence on the sheet in front of you. It is important that you make a guess at the word if you are unsure of your answer.

You are also going to be asked to rate the speaker on a scale of 1-7 for acceptability. 1 represents speech which was least acceptable and 7 speech which was highly acceptable. In making your judgments about the speakers you are about to hear, give careful consideration to the attributes of pitch, rate, understandability, and voice quality. In other words, is the voice pleasing to listen to, or does it cause you some discomfort as a listener?

You will also be asked to record your subjective opinion of the sex of the speaker.
Appendix J-continued

Speakers

You are about to participate in an experiment related to the understanding of speech signals produced by an artificial larynx.

You will be asked to produce some sentences using an artificial larynx. We will spend approximately 5 minutes giving you instructions related to the production of speech with the artificial larynx. During this practice session you will be expected to maintain a constant intensity level (loudness) while producing the sentences. A VU meter will be provided as an aid to help you maintain this constant intensity (loudness) level throughout the experiment.

You will be provided with a series of word lists consisting of 50 words each. You will produce each word at the end of the carrier phrase, "You will say _____". Following each of these sentence productions you will pause for 3 seconds before starting the next sentence. I will provide examples of this procedure before we begin.

Please feel free to signal with your hand if you would like to rest.

Do you have any concerns or questions?

I have read these instructions, and I have had all questions answered to my satisfaction.

Name _____________________________

Date ______________________________
Appendix K

Informed Consent Release Form

1. I ________________ freely and voluntarily consent to participate in the experiment described on the attached page.

2. I also understand that I may withdraw from this experiment at any time and that my participation or withdrawal will in no way affect my standing as a student in this university or as a consumer of its clinical offerings.

3. I understand that I will not be exposed to any experimental procedures which would in any way be detrimental to my physical or psychological well being.

4. I understand that other individuals will be participating in the experiment with me. However, I also understand that none of my responses will in any way be associated with me or with my name.

5. I engage in this study freely, without monetary payment and with no other contingencies being placed upon my participation. I also understand that I will not benefit personally from the results of the experiment.

6. I understand that I have had and will have the opportunity to ask questions about the nature and purpose of the study, and I understand that upon the completion of this study at my request, I can obtain additional explanation about this study and its implications.

Date _______________________________ signed _______________________________ witness

____________________________ witness

For additional information contact John M. Hanley Ph.D. (383-0963) or Sandra Merritt (388-3547).
Appendix L
Listener's Response Sheet

Date ____________________
Tape ____________________

1. ______________________ 26. ______________________
2. ______________________ 27. ______________________
3. ______________________ 28. ______________________
4. ______________________ 29. ______________________
5. ______________________ 30. ______________________
6. ______________________ 31. ______________________
7. ______________________ 32. ______________________
8. ______________________ 33. ______________________
9. ______________________ 34. ______________________
10. ______________________ 35. ______________________
11. ______________________ 36. ______________________
12. ______________________ 37. ______________________
13. ______________________ 38. ______________________
14. ______________________ 39. ______________________
15. ______________________ 40. ______________________
16. ______________________ 41. ______________________
17. ______________________ 42. ______________________
18. ______________________ 43. ______________________
19. ______________________ 44. ______________________
20. ______________________ 45. ______________________
21. ______________________ 46. ______________________
22. ______________________ 47. ______________________
23. ______________________ 48. ______________________
24. ______________________ 49. ______________________
25. ______________________ 50. ______________________

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Appendix M

Matrix of Percentages of Total Phoneme Error Substitutions by Speaker. Percentages of Frequencies
A(94Hz.), B(120Hz.), C(160Hz.), D(212Hz.), E(245Hz.)

Speaker 1

<table>
<thead>
<tr>
<th></th>
<th>p</th>
<th>t</th>
<th>k</th>
<th>b</th>
<th>d</th>
<th>g</th>
<th>glide</th>
<th>fric</th>
<th>omis</th>
<th>m/n</th>
</tr>
</thead>
<tbody>
<tr>
<td>p</td>
<td></td>
<td></td>
<td></td>
<td>11,14, 28,26,22</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>t</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>k</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>28,27,27</td>
<td>0.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>d</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>g</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0,20,20</td>
<td>0.60</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>glide</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>fric</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>50,17,0</td>
<td>61,17,20</td>
<td></td>
</tr>
<tr>
<td>omis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0,33</td>
<td>8,11</td>
<td></td>
</tr>
<tr>
<td>m/n</td>
<td>78,11,0</td>
<td>11,1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>h</td>
<td>16,8,24</td>
<td>28,24</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>r</td>
<td>7,14,7</td>
<td>29,43</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>added</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Speaker 2</td>
<td>m/n</td>
<td>fric</td>
<td>glide</td>
<td>omis</td>
<td>t</td>
<td>k</td>
<td>b</td>
<td>d</td>
<td>g</td>
<td>e</td>
</tr>
<tr>
<td>-----------</td>
<td>-----</td>
<td>------</td>
<td>-------</td>
<td>------</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
</tr>
<tr>
<td></td>
<td>20,18,20</td>
<td></td>
<td></td>
<td></td>
<td>8,17,8</td>
<td>3,13,5</td>
<td>15,15,7</td>
<td>61,0,7</td>
<td>25,15,50</td>
<td>18,15,20</td>
</tr>
<tr>
<td></td>
<td>26,18,20</td>
<td></td>
<td></td>
<td></td>
<td>8,17,8</td>
<td>3,13,5</td>
<td>15,15,7</td>
<td>61,0,7</td>
<td>25,15,50</td>
<td>18,15,20</td>
</tr>
<tr>
<td></td>
<td>6,21,28</td>
<td></td>
<td></td>
<td></td>
<td>33,25,29</td>
<td>23,17,39</td>
<td>33,25,29</td>
<td>23,17,39</td>
<td>33,25,29</td>
<td>23,17,39</td>
</tr>
<tr>
<td></td>
<td>7,22,31</td>
<td></td>
<td></td>
<td></td>
<td>15,5,20</td>
<td>56,15,18</td>
<td>10,15,20</td>
<td>56,15,18</td>
<td>10,15,20</td>
<td>56,15,18</td>
</tr>
<tr>
<td></td>
<td>18,17</td>
<td></td>
<td></td>
<td></td>
<td>29,2,36,2</td>
<td>34,20,50</td>
<td>34,20,50</td>
<td>34,20,50</td>
<td>34,20,50</td>
<td>34,20,50</td>
</tr>
</tbody>
</table>

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Appendix M — Continued

Speaker 3

<table>
<thead>
<tr>
<th></th>
<th>p</th>
<th>t</th>
<th>k</th>
<th>b</th>
<th>d</th>
<th>g</th>
<th>glide</th>
<th>fric</th>
<th>omis</th>
<th>m/n</th>
</tr>
</thead>
<tbody>
<tr>
<td>p</td>
<td>38.9</td>
<td>21,17,13</td>
<td>14.27,4</td>
<td>0,0,0,100.0,39.6</td>
<td>0,14,24,20,13,35</td>
<td>0,29,14,57</td>
<td>11,17,6,47,0,0,0,23,32</td>
<td>12,37,6</td>
<td>12,37,6</td>
<td></td>
</tr>
<tr>
<td>t</td>
<td>0.36,0</td>
<td>22.7</td>
<td>17.26</td>
<td>11,17,6,47,0,0,0,23,32</td>
<td>12,37,6</td>
<td>23,32</td>
<td>14.27,4</td>
<td>12,37,6</td>
<td>12,37,6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>45.18</td>
<td>0,14</td>
<td>24</td>
<td>11,17,6,47,0,0,0,23,32</td>
<td>12,37,6</td>
<td>23,32</td>
<td>14.27,4</td>
<td>12,37,6</td>
<td>12,37,6</td>
<td></td>
</tr>
<tr>
<td>k</td>
<td>14.9,14</td>
<td>3.12,26,18</td>
<td>26.53,6,16,16,24</td>
<td>20.10,10,10,50</td>
<td>15,38,27</td>
<td>29,0</td>
<td>20,10,10,50</td>
<td>15,38,27</td>
<td>29,0</td>
<td></td>
</tr>
<tr>
<td>b</td>
<td>36.27</td>
<td>3.12,26,18</td>
<td>26.53,6,16,16,24</td>
<td>20.10,10,10,50</td>
<td>15,38,27</td>
<td>29,0</td>
<td>20,10,10,50</td>
<td>15,38,27</td>
<td>29,0</td>
<td></td>
</tr>
<tr>
<td>d</td>
<td>38,38,23</td>
<td>20.10,10,10,50</td>
<td>15,38,27</td>
<td>29,0</td>
<td>20,10,10,50</td>
<td>15,38,27</td>
<td>29,0</td>
<td>20,10,10,50</td>
<td>15,38,27</td>
<td></td>
</tr>
<tr>
<td>g</td>
<td>8,8</td>
<td>20,10,10,50</td>
<td>15,38,27</td>
<td>29,0</td>
<td>20,10,10,50</td>
<td>15,38,27</td>
<td>29,0</td>
<td>20,10,10,50</td>
<td>15,38,27</td>
<td></td>
</tr>
<tr>
<td>glide</td>
<td>7,14,7</td>
<td>50,21</td>
<td>7,14,7</td>
<td>50,21</td>
<td>7,14,7</td>
<td>50,21</td>
<td>7,14,7</td>
<td>50,21</td>
<td>7,14,7</td>
<td></td>
</tr>
<tr>
<td>fric</td>
<td>0.21</td>
<td>6.47</td>
<td>8.24,19</td>
<td>38,38,23,24,25</td>
<td>38,38,23,24,25</td>
<td>38,38,23,24,25</td>
<td>38,38,23,24,25</td>
<td>38,38,23,24,25</td>
<td>38,38,23,24,25</td>
<td></td>
</tr>
<tr>
<td>m/n</td>
<td>14,24,0</td>
<td>27,10,13</td>
<td>27,10,13</td>
<td>27,10,13</td>
<td>27,10,13</td>
<td>27,10,13</td>
<td>27,10,13</td>
<td>27,10,13</td>
<td>27,10,13</td>
<td></td>
</tr>
<tr>
<td>h</td>
<td>33,29</td>
<td>7,43</td>
<td>7,43</td>
<td>7,43</td>
<td>7,43</td>
<td>7,43</td>
<td>7,43</td>
<td>7,43</td>
<td>7,43</td>
<td></td>
</tr>
<tr>
<td>r</td>
<td>36.7,18</td>
<td>21,18</td>
<td>21,18</td>
<td>21,18</td>
<td>21,18</td>
<td>21,18</td>
<td>21,18</td>
<td>21,18</td>
<td>21,18</td>
<td></td>
</tr>
<tr>
<td>added</td>
<td>39,13,3</td>
<td>12,44,20</td>
<td>12,44,20</td>
<td>12,44,20</td>
<td>12,44,20</td>
<td>12,44,20</td>
<td>12,44,20</td>
<td>12,44,20</td>
<td>12,44,20</td>
<td></td>
</tr>
<tr>
<td></td>
<td>10,32</td>
<td>8.16</td>
<td>8.16</td>
<td>8.16</td>
<td>8.16</td>
<td>8.16</td>
<td>8.16</td>
<td>8.16</td>
<td>8.16</td>
<td></td>
</tr>
</tbody>
</table>
BIBLIOGRAPHY


Lafon, J., & Cornut, G. Etude de la formation impulsionnelle de la voix et de la parole. Folio Phoniatrica, 1960, 12, 176-188.


