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.
Trudeau, Michael David

FACTORS AFFECTING THE ACCEPTABILITY OF NORMAL LARYNGEAL, ESOPHAGEAL AND PULMONARY ASSISTED ALARYNGEAL VOICE

The Ohio State University

Ph.D. 1982

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

In all cases this material has been filmed in the best possible way from the available copy. Problems encountered with this document have been identified here with a check mark √.

1. Glossy photographs or pages _____
2. Colored illustrations, paper or print _____
3. Photographs with dark background _____
4. Illustrations are poor copy _____
5. Pages with black marks, not original copy _____
6. Print shows through as there is text on both sides of page _____
7. Indistinct, broken or small print on several pages __
8. Print exceeds margin requirements _____
9. Tightly bound copy with print lost in spine ______
10. Computer printout pages with indistinct print _____
11. Page(s) ___________ lacking when material received, and not available from school or author.
12. Page(s) ___________ seem to be missing in numbering only as text follows.
13. Two pages numbered ___________. Text follows.
14. Curling and wrinkled pages _____
15. Other ____________________________

University Microfilms International
FACTORs AFFECTING THE ACCEPTABILITY OF
NORMAL LARYNGEAL, ESOPHAGEAL AND
PULMONARY ASSISTED ALARYNGEAL VOICE

DISSERTATION

Presented in Partial Fulfillment of the Requirements for
the Degree of Doctor of Philosophy in the Graduate
School of The Ohio State University

By

Michael David Trudeau, B.A., M.A.

* * * * *

The Ohio State University

1982

Reading Committee:               Approved By
Edward J. Hardick
Sheila M. Goff
Jane E. Jarrow

Edward J. Hardick
Advisor
Department of Communication
To SETH
who was always there
to remind me
to burn the midnight oil,

To ARIEL
who was always there
to remind me
that the early bird catches the worm,

To LINDA
who was
always there.
ACKNOWLEDGEMENTS

The amount of support and encouragement offered to this investigator was both gratifying and humbling. I was fortunate in this effort to benefit from the wisdom of two advisors, Sheila M. Goff, Ph.D. and Edward J. Hardick, Ph.D. To both I extend my sincerest gratitude. Each gave the darkest clouds a silver lining. To the remaining members of my committee, I also owe a debt of deep appreciation: to John J. Kennedy, Ph.D. for his invaluable assistance in the statistical design; to Jane E. Jarrow, Ph.D. for her probing questions, editorial scalpel, and listening ability; and to David E. Schuller, M.D. for his help in identifying subjects.

My indebtedness extends well beyond my committee. Claude Lambert, M.A. was a constant source of technical information and assistance. Tammy Thompson, M.A., lent a critical ear in judging speaker proficiency. Charles W. Nixon, Ph.D., Richard McKinley, M.S., and the staff of the Air Force Aerospace Medical Research Laboratory were invaluable in their technical and theoretical help. Myron Shapiro, M.D. allowed me the courtesy of recording three of his patients. Cindy Spillers, M.A. recorded 20% of the alaryngeal speakers in this study. The faculty and staff of the Department of Otolaryngology, The Ohio State University Hospitals were intrinsic in identifying potential subjects. Finally my profound thanks are due to the laryngeal and alaryngeal speakers who kindly volunteered their time and voices and without whom there could have been no study.
VITA

November 17, 1949...... Born, New Bedford, Massachusetts

1972....................... B.A., Slavic Languages and Literature, The Ohio State University, Columbus, Ohio.


1978-1979.............. Speech/Language Pathologist, Delaware Speech and Hearing Center, Delaware, Ohio.

1979....................... M.A., Speech Pathology, The Ohio State University, Columbus, Ohio.

1979-1980.............. Assistant Director, Delaware Speech and Hearing Center, Delaware, Ohio.


1982....................... Ph.D., Speech Pathology, The Ohio State University, Columbus, Ohio.

PRESENTATIONS


Miniseminar: "Tracheoesophageal Puncture for Voice Restoration Following Total Laryngectomy", Ohio Speech and Hearing Association Convention, Cincinnati, Ohio, April, 1981.

GRANT

Graduate Student Alumni Research Award, December, 1981 to September, 1982.
FIELDS OF STUDY

Major Field: Speech Pathology

Studies in Neuropsychology. Professor Donald Meyers

Studies in Experimental Design and Statistical Methodology. Professor John J. Kennedy

Studies in Surgical-Prosthetic Alaryngeal Voice Restoration. Assistant Professor Jane E. Jarrow
# TABLE OF CONTENTS

DEDICATION .............................................................................................................. ii  
ACKNOWLEDGEMENTS .......................................................................................... iii  
VITA ................................................................................................................................ iv  
LIST OF TABLES ................................................................................................... viii  
LIST OF FIGURES ................................................................................................. xi  

Chapter

I. INTRODUCTION ................................................................................................. 1  
   Statement of the Problem .................................................................................. 2  
   Purpose.............................................................................................................. 3  
   Hypotheses....................................................................................................... 3  
   Definition of Terms........................................................................................... 4  
   Organization of Study....................................................................................... 6  

II. REVIEW OF THE LITERATURE ....................................................................... 7  
   Source-Filter Theory......................................................................................... 7  
   Frequency Related Measures........................................................................... 11  
   Intensity Related............................................................................................... 20  
   Temporal Measures.......................................................................................... 25  
   Perceptual Measures....................................................................................... 29  
   Summary.......................................................................................................... 36  

III. PROCEDURES .................................................................................................. 38  
   The Problem..................................................................................................... 38  
   Speaking Subjects............................................................................................. 38  
   Recording Instrumentation............................................................................... 40  
   Physical Measures Methodology and Equipment ........................................... 46  
   Equal-Appearing Interval Scale........................................................................ 53  
   Summary.......................................................................................................... 60  

vii
## IV. RESULTS AND DISCUSSION

- **Acceptability**
- **Frequency**
- **Intensity**
- **Temporal Measures**
- **Multiple Regression Analysis**

## V. SUMMARY AND CONCLUSIONS

**Conclusion**

## BIBLIOGRAPHY

## APPENDICES

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Form for Speech Pathologist's Evaluation of Speaker Vocal Proficiency</td>
<td>98</td>
</tr>
<tr>
<td>B. Participant Information Sheet/Consent Form</td>
<td>102</td>
</tr>
<tr>
<td>C. Speaker Background Information Questionnaire</td>
<td>104</td>
</tr>
<tr>
<td>D. Rainbow Passage</td>
<td>108</td>
</tr>
<tr>
<td>E. Eleven &quot;H_D&quot; Sentences</td>
<td>110</td>
</tr>
<tr>
<td>F. Recording Session Procedures</td>
<td>112</td>
</tr>
<tr>
<td>G. Verbal Directions to Listener-Judges</td>
<td>116</td>
</tr>
<tr>
<td>H. Acceptability Form for Scale Construction</td>
<td>118</td>
</tr>
<tr>
<td>I. Acceptability Form for Assessing Experimental Voices</td>
<td>120</td>
</tr>
</tbody>
</table>
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-1. Representative Formant 1-3 Values for normal, Esophageal and Tracheoesophageal Male Speakers</td>
<td>17</td>
</tr>
<tr>
<td>2-2. Representative Formant 1-3 Values for Normal and Esophageal Female Speakers</td>
<td>19</td>
</tr>
<tr>
<td>2-3. Intensity Characteristics of Esophageal, Tracheoesophageal and Normal Male Speakers</td>
<td>23</td>
</tr>
<tr>
<td>2-4. Intensity Control of Esophageal, T-E, and Normal Male Speakers as Demonstrated in Intensity Standard Deviation, Vocal Shimmer and Shimmer Standard Deviation</td>
<td>23</td>
</tr>
<tr>
<td>2-5. Comparison of Mean Reading Rate and Mean Percent of Pause Time in Reading for Esophageal and Paa Speakers</td>
<td>30</td>
</tr>
<tr>
<td>2-6. Summary of Intelligibility Scores for Normal, Esophageal and T-E Speakers in Quiet Speaking and Listening Conditions</td>
<td>32</td>
</tr>
<tr>
<td>3-1. Characteristics of Listeners in Construction of Acceptability Scale</td>
<td>56</td>
</tr>
<tr>
<td>4-1. Means and Standard Deviations of Acceptability by Voice Type and Sex of Speaker</td>
<td>62</td>
</tr>
<tr>
<td>4-2. Repeated Measures ANOVA Summary Table for Acceptability of Voice by Voice Type and Sex of Speaker</td>
<td>62</td>
</tr>
<tr>
<td>4-3. Means and Standard Deviations of $F_2:F_1$ for the Quantal Vowels, for Range of $f_o$ for the Quantal Vowels, for $f_o /a/$ for the Quantal Vowels by Voice and Sex</td>
<td>67</td>
</tr>
<tr>
<td>4-4. MANOVA Summary Table for Formant Ratios and Range of $f_o$ for the Quantal Vowels and for $f_o /a/$</td>
<td>68</td>
</tr>
<tr>
<td>4-5. Summary of Univariate F-Tests for $F_2:F_1$ of /i,a,u/ and Range of $f_o$ across /i,a,u/ and $f_o /a/$. D.F.(2,39)</td>
<td>68</td>
</tr>
<tr>
<td>4-6. Weighted Means for $F_2:F_1$ of /i,a,u/, Range of $f_o$ across /i,a,u/ and $f_o$ by Voice and Sex</td>
<td>69</td>
</tr>
</tbody>
</table>
4-7. Standardized Weights and Structure Coefficients for 
\( F_2,F_1 \) of /i,a,u/, Range of \( f_o \) across /i,a,u/ and \( f_o /a/ 
by Voice and Sex ................................................................. 69

4-8. Means and Standard Deviations of the First Two For­
mants of the Quantal Vowels by Voice Type and Sex ....... 70

4-9. MANOVA Summary Table for the First Two Formants 
of Quantal Vowels by Sex ...................................................... 71

4-10. Summary of Univariate F-Tests for the First Two For­
mants of Quantal Vowels by Sex. D.F.(1,39) ....................... 71

4-11. MANOVA Summary Table for First Two Formants of 
Quantal Vowels by Voice Type ............................................ 71

4-12. Summary of Univariate F-Tests for First Two For­
mants of Quantal Vowels by Voice Type ............................... 73

4-13. Means and Standard Deviations of Softest Phonated 
/a/ (XS), loudest comfortable /a/ (XL), least to most 
tense /a/ (LTM) and Range from XS to XL (XR) by 
Voice Type and Sex ............................................................... 74

4-14. MANOVA Summary Table for Effects of Intensity Fac­
tors by Voice and Sex ........................................................... 75

4-15. Standardized Weights and Structure Coefficients for 
XS, XL, LTM, and XR by Voice Type and Sex ...................... 75

4-16. Weighted Means (WX) of XS, XL, LTM, and XR by 
Voice Type and Sex ............................................................. 75

4-17. Summary of Univariate F-Tests for Intensity Factors 
by Voice Type and Sex .......................................................... 76

4-18. Means and Standard Deviations of Temporal Measures 
by Sex and Voice ................................................................. 79

4-19. MANOVA Summary Table for Temporal Measures by 
Voice and Sex ....................................................................... 79

4-20. Summary of Univariate F-Tests for Temporal Measures 
by Voice And Sex. D.F.(2,39) ...................................................... 80

4-21. Weighted Means for Temporal Measures by Voice and 
Sex ....................................................................................... 80

4-22. Standardized Weights and Structure Coefficients for 
Temporal Measures by Voice and Sex ................................. 80

4-23. MANOVA Summary Table for Temporal Measures by Sex... 81
4-24. Summary of Univariate F-Test for Temporal Measures by Sex. D.F.(1,39) ................................................................. 81

4-25. Table of Weighted Means for Temporal Measures by Sex. ................................................................. 81

4-26. Standardized Weights and Structure Coefficients for Temporal Measures by Sex ................................................................. 81

4-27. MANOVA Summary Table for Temporal Measures by Voice Type ........................................................................... 82

4-28. Weighted Means of Temporal Measures by Voice Type ........................................................................... 82

4-29. Standardized Weights and Structure Coefficients for Temporal Measures by Voice ........................................................................... 82

4-30. Multiple Regression Models of the Correlation of the Three Dependent Variables (XL, f/a/, Rf), of the Two Composite Variables (Phonatory Proficiency and Formant Elevation), and Location and Sex of Speaker with Vocal Acceptability (F-Tests compare Adjacent Models) ........................................................................... 85

4-31. Multiple Regression Models of Variable Correlation with Acceptability Scores with F-Test comparing Adjacent Models ........................................................................... 88
LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4-1. Graph of Y Score Means for Intensity Data</td>
<td>78</td>
</tr>
</tbody>
</table>

xi
FACTORS AFFECTING THE ACCEPTABILITY OF NORMAL LARYNGEAL, ESOPHAGEAL AND PULMONARY ASSISTED ALARYNGEAL VOICE

By

Michael D. Trudeau, Ph.D.
The Ohio State University

Edward J. Hardick, Ph.D., Advisor

45 age-matched speakers were divided into three equal groups (10 males, 5 females) based on type of voice: normal, laryngeal voice, esophageal voice, and pulmonary assisted alaryngeal (paa) voice. Audio recordings of each subject's voice were made during the performance of various tasks. The resultant voice samples were judged for acceptability by 25 naive listeners trained in use of a 5 point equal-appearing interval scale. Acoustic analyses of the voice samples were also performed. A number of statistical procedures were applied revealing: 1. Paa and esophageal speakers differed significantly from laryngeal speakers but not from each other in acceptability; 2. The primary acoustic characteristics contributing to acceptability were intensity of loudest comfortable /a/, vocal $f_0$, temporal proficiency in producing speech, formant elevation, and range of $f_0$ across the quantal vowels /i,a,u/. The author concluded that the results provided support for choosing prosthetic voice restoration following laryngectomy as a viable alternative to esophageal voice training.
CHAPTER I
Introduction

For almost 100 years surgical or surgical-prosthetic approaches to alaryngeal voice restoration (the utilization of existing anatomical structures as a neoglottis for pulmonary air) have been reported in the professional literature (Park, 1866; Blom and Singer, 1979). This theme recurs for two reasons: first, other forms of alaryngeal voice (artificial larynx or esophageal phonation) display inherent physical or perceptual limitations (Snidecor and Curry, 1959; McCroskey and Mulligan, 1963; and Hartman, 1979). Second, while surgical-prosthetic management may alleviate many of the limitations of the alternative methods, complications with aspiration via the shunt and/or stenosis of the shunt have proscribed wide acceptance of surgical-prosthetic techniques (Blom and Singer, 1979; 1980).

Recently at least three highly similar solutions to the problems of aspiration and stenosis have appeared: the Blom-Singer Voice Prosthesis (BSVP), the Panje Voice Button (PVB), and the Shapiro Procedure (SP). In the former two methods a tracheoesophageal (t-e) puncture or incision is used to provide an airway connecting the tracheostoma at the level of, or just below, the pharyngoesophageal (p-e) segment. For these two procedures the assumed site of vibration is the p-e segment. In the latter method, the esophagus is bypassed and a tracheohypopharyngeal (thp) puncture produces an airway connecting
the stoma to the hypopharynx near the base of the tongue. For SP speech the assumed site of vibration is the pharyngeal mucosa. All three procedures are similar in that placement of a valved, silastic prosthesis within the surgical shunt serves to avert aspiration and to maintain patency. Occlusion of the stoma on exhalation diverts the pulmonary air stream through the prosthesis. The air stream activates the respective site of vibration and phonation results. Cessation of expiration terminates phonation (Blom and Singer, 1979; Blom, 1981; Shapiro, 1982). Due to the strong similarity among these three methods, all will be considered under the rubric of pulmonary assisted alaryngeal (paa) voice.

**Statement of the Problem**

While these three devices appear to pose satisfactory solutions to the problems of aspiration and stenosis and the intensity of the vocal product is strengthened in comparison to esophageal voice, the perceptual effects of paa voice have not been explored. Two related questions arise: 1) Is paa voice more pleasing perceptually than more traditional esophageal voice? 2) If there are perceptual differences, what physical characteristics of the voice produced the perceptual phenomenon?

Robbins, et al.(1981) performed extensive acoustic analyses of the voices of esophageal speakers. A linear discriminant analysis was applied to the acoustic analyses data in order to determine if the three groups could be differentiated from each other and, if so, by what factors. The results indicated that the linear discriminant analysis could differentiate the groups and account for 100% of the variance. The most discriminating variables in order of strength were median intensity during reading, shimmer ratio, maximum phonation time, percent of pause time in oral reading, and jitter ratio. These
findings suggest that paa voice may be perceptually superior to esophageal voice and inferior to normal laryngeal voice. However, Robbins, et.al. did not obtain perceptual measures; therefore a conclusive statement concerning the relationship of the physical measures to the perceptual effects was not possible.

A clinical question also arises from the recent advances in paa voice and from the results of the Robbins, et.al. study: to wit, are the measures of vocal performance appropriate for documenting accountability in the communicative management of the esophageal voice client still viable for documenting the acquisition of paa voice? While Robbins, et.al. supplied five discriminating variables, only one of these, maximum phonation time, can be derived readily in a clinical setting. The remaining four variables require the instruments of a well equipped speech science laboratory for measurement.

Purpose

The purpose of this study is to describe and relate the physical characteristics of paa voice to similar characteristics of esophageal and normal voice and to identify how these physical characteristics affect the perception of vocal acceptability (a listener's judgement of how pleasant or unpleasant a voice is). A secondary purpose is to attempt to identify clinically obtainable measures of vocal performance which may serve as useful indices of proficiency in developing paa voice.

Hypotheses

To compass the purpose of this study, four hypotheses are proposed:

1. In terms of acceptability of voice, no significant differences exist among normal laryngeal, esophageal, and paa speakers.
II. In terms of vocal acoustic characteristics, no significant differences exist among normal laryngeal, esophageal, and paa speakers.

III. In terms of proficiency of voice production, no significant differences exist among normal laryngeal, esophageal, and paa speakers.

IV. No combination of measures of vocal proficiency and vocal acoustic characteristics can predict reliably a speaker's rating of vocal acceptability.

Definition of Terms:

In order to ensure a common frame of reference, the following list of terms used in the study of normal and deviant voice is provided.

1. Acceptability -- the perceptual judgement of how pleasant or unpleasant a voice is.

2. Alaryngeal -- without a larynx.

3. Blom-Singer Voice Prosthesis -- a silastic valved tube used to prevent aspiration and shunt stenosis in tracheoesophageal speakers. Available in lengths varying from 2.2. to 4.3 cm (abbr. BSVP).

4. Charge -- the act of forcing intraoral air into the esophagus. This act is a necessary antecedent to esophageal phonation.

5. Electro-larynx -- an electro-mechanical device used to produce a pseudo-glottal tone in order to restore vocal communication for the laryngectomized.

6. Esophageal voice -- a form of alaryngeal voice produced by forcing air into the esophagus and releasing that air to set the pharyngo-esophageal segment into vibration.

7. Formant -- an area of maximum energy transfer due to the resonant characteristics of the vocal tract. Usually identified by a midpoint frequency (abbr. $F_1$, $F_2$, ..., $F_n$).

8. Fundamental frequency -- the rate of vibration of a phonatory body. Usually expressed in cycles per second (Hertz, Hz) (abbr. $f_0$).
9. **Intelligibility** -- the degree of agreement between the listener's perception of what was said and the speaker's production of what was said.

10. **Jitter** -- cycle-to-cycle variability of the period of the vocal fundamental frequency.

11. **Panje Voice Button** -- a silastic valved tube used to prevent aspiration and shunt stenosis in tracheoesophageal speakers (abbr. PVB). Unlike BSVP, this device comes in only two lengths.

12. **Pharyngoesophageal segment** -- a group of muscles which acts as a valve between the esophagus and pharynx. The assumed site of phonation in tracheoesophageal voice (abbr. p-e).

13. **Pulmonary-assisted alaryngeal voice** -- alaryngeal voice produced by diverting the pulmonary air stream into the esophagus or pharynx, where phonation occurs (abbr. paa).

14. **Shapiro Procedure** -- a silastic stoma vent with a 4.3 cm BSVP placed vertically from the top. Used to prevent aspiration and shunt stenosis in tracheohypopharyngeal speakers (abbr. SP).

15. **Shimmer** -- cycle-to-cycle variability of the intensity of the vocal fundamental frequency.

16. **Staffieri Procedure** -- a surgical method in which the trachea communicates with either the esophagus or hypopharynx via a surgical incision. Patency and elasticity of the incision are preserved by lining the incision with mucosal tissue, rather than by filling the shunt with a valved prosthesis.

17. **Stoma noise** -- audible air turbulence created by forceful exhalation of air through the tracheostoma.

18. **Tracheoesophageal voice** -- voice produced by diverting pulmonary air from the trachea to the esophagus (abbr. t-e).
19. Tracheohypopharyngeal voice - - voice produced by diverting pulmonary air from the trachea to the hypopharynx (abbr. thp).

20. Tracheostoma - - the opening at the anterior base of the neck, through which respiration occurs following total laryngectomy.

**Organization of the Study**

Within this chapter a statement of the problem, the purpose of this study, four research hypotheses, and definitions of relevant terms have been presented. The subsequent chapter provides a review of the literature divided into the following topics: source-filter theory, frequency related measures of voice, intensity related measures of voice, temporal measures of voice, and perceptual measure of voice. Chapter III describes the procedures used in this study. Major topics in that chapter are methodology for obtaining voice samples, group characteristics of speakers, instrumentation utilized in analysis samples, group characteristics of speakers, instrumentation utilized in analysis of the voice samples, construction and use of acceptability scale and group characteristics of listeners. Chapter IV presents the results of this study and a discussion. Chapter V summarizes this study and suggests questions for future research.
CHAPTER II
Review of the Literature

The purpose of this study is to describe the vocal factors by which esophageal, pulmonary assisted alaryngeal (paa), and normal speakers differ or are similar and to relate these factors to the perception of vocal acceptability. The contribution of this chapter to that purpose is to provide a conceptual framework for addressing research in voice production, and to describe the results of earlier investigations into this area. Chapter II has been divided into six sections: a) source-filter theory, b) frequency related measures of voice, c) intensity related measures of voice, d) temporal measures of voice, e) perceptual measures of voice, and f) a summarization of the measures employed in this study.

Source-Filter Theory

Investigation into the nature of human voice production, either normal or deviant, generally arises from the same conceptual framework, the acoustics theory of human communication. This theory describes communication as a sequence of events involving a speaker, a listener, and an acoustic medium. The speaker formulates a message and transmits that message to the listener by effecting systematic alterations of the acoustic environment with the vocal mechanism. The listener receives these acoustic signals and decodes them to derive meaning (Denes and Pinson, 1972; Dudley, 1940). Within this
model the production of voice assumes an integral role as the means by which the acoustic environment is disturbed by the speaker to affect the listener. The vocal signal is the bridge that connects the central nervous systems of the speaker and the listener. The proper study of the human voice, whether normal or deviant, therefore, must encompass two broad sets of occurrences: the events in the physical or acoustic environment and the events in the neural or perceptual environment (Lieberman, 1977).

While the acoustic theory provides a broad perspective, a narrower view of voice production may be gained through one aspect of the acoustic theory, the source-filter theory of voice (Lieberman, 1977; Michel and Wendahl, 1971). This theory requires three components: an energy supply, usually the pulmonary airstream; a source of vibration, normally the vocal folds; and a filtering mechanism, the resonant characteristics of the vocal tract.

The foundations of the source-filter theory were set in the mid and late nineteenth century as researchers sought to explain how vowels were differentiated. Initially Wheatstone postulated the harmonic theory of vowel production, Willis countered with the air-puff theory, and Helmholtz discerned the partial value of each position and produced a consolidated explanation (Lindsay, 1966; Fletcher, 1953). Wheatstone described the formation of vowels as a result of a fundamental tone created by vocal fold vibration and of harmonics of the glottal tone which were modified by the vocal tract. Thus, Wheatstone emphasized the contribution of the phonatory source. Willis' theory diverged from Wheatstone's position by relegating the vocal folds to the role of a valve which released the pulmonary airstream to energize the inherent resonant characteristics of the pharyngeal, oral, and nasal cavities. Willis, therefore, placed greater import on vocal tract resonance than did
Wheatstone. Helmholtz combined the two theories to furnish a single, coherent, and still valid theory in which the glottal tone served as the initial source of acoustic disturbance and the resonances of the vocal tract cavities served to filter and modify the glottal tone. These resonances of the vocal tract which are intrinsic in vowel discrimination are designated as formants. Peterson and Barney (1952) identified the first three formants for each of the ten vowels /i, ɛ, ι, a, η, u, ɔ, ɔ, a/. From their work, the relationship of the formants in defining vowels became apparent: the ratio of the first two formants to each other is the basic acoustic cue for identifying the first nine (non-retroflexed) vowels.

The ratio of $F_2:F_1$ is not endlessly variable; rather the physical limitations of the normal vocal tract serve to set boundaries on this ratio. The vowels /i, a, u/ appear to delineate or occur at these boundaries and the production of the remaining vowels is determined by the location of /i, a, u/. Because of their relative prominence in defining the possible range of vowels, these three vowels have been labelled as the quantal vowels (Lieberman, 1977). Since the quantal vowels fall at the limits of the $F_2:F_1$ ratio for normal speakers, they are of interest in investigating the effects of total laryngectomy, since this surgery alters the resonant characteristics of the vocal tract.

With respect to the source-filter theory, total laryngectomy obviously has a drastic effect on the source of phonation: the vocal folds are no longer present to vibrate. Secondary effects are present, however, which influence the other two components of the theory as well. The primary purpose of the larynx is to act as a valve which protects the lower respiratory system from the entry of foreign material (Zemlin, 1968). To prevent aspiration following laryngectomy, the respiratory pathway must be diverted from the pharynx to
exit the body via a tracheostoma at the base of the neck. The pulmonary air-stream cannot pass through the vocal tract and, therefore, without surgical modification it cannot act as the energy supply for voice. Additionally, the larynx not only functioned to create the glottal tone, but also altered the resonant features of the vocal tract by raising or lowering during speech (Lieberman, 1977; Zemlin, 1968). Removal of the larynx changes the resonant characteristics of the vocal tract, probably by shortening the length of the resonating tube (Sisty and Weinberg, 1972). In summary, total laryngectomy deprives the patient of the phonatory source, diverts the power supply for voice from the vocal tract, and alters the filtering function of the vocal mechanism. These changes in the main components of the source filter theory can be shown to affect the vocal production of laryngectomized speakers; however, the conceptual framework provided by this theory still may be applied, allowing direct comparisons of alaryngeal and normal, laryngeal voice.

Three physical parameters are useful in describing sound: frequency, intensity, and time (Fry, 1977). These three have perceptual correlates of pitch, loudness, and duration, respectively (Fletcher, 1953). A fourth, less well defined perceptual parameter also is applied to voice: quality. Where the three former perceptual parameters may be defined as the auditory sensation of their respective, physical counterparts, quality has no direct physical correlate and appears to be affected by all three physical parameters (Michel and Wendahl, 1971).
Frequency Related Measures

The frequency related characteristics of voice are measurable in a variety of ways. Those measures most commonly reported in the literature are fundamental frequency, and formant frequencies for at least formants one and two.

Fundamental frequency ($f_o$) has been reported as either the mean or median frequency of vibration of the glottal tone (Horii, 1975). Substantively, the reporting of mean or median values appears to make little difference. Horii (1975) compared mean and median values of $f_o$ derived from oral reading of the "Rainbow Passage" for the same speakers. Mean $f_o$ was reported as 112.5 Hz in a range of 84-151 Hz, and median $f_o$ was reported as 110.7 Hz in a range of 82-150 Hz. The group of subjects used in this study were not homogeneous with respect to sex or age. Both sex and age are factors which influence $f_o$.

For normal, young, male adults, Hollien and Jackson (1973) reported a mean $f_o$ in oral reading of 129.4 Hz with a range of 92.6 to 178.1 Hz. The mean age of this sample was 20.3 years. Mysak (1959) reported mean $f_o$ for three groups of older male adults. The mean age of each group was 47.9 years, 73.3 years, and 85.0 years. Mean $f_o$ for each group was 113.2 Hz, 124.3 Hz and 141.0 Hz. In a similar study with aged women, McGlone and Hollien (1963) found mean $f_o$ for subjects aged 67-79 years to be 196.6 Hz, and for subjects aged 80-94 years to be 199.8 Hz. For females with a mean age of 24.6 years, Stoicheff (1981) reported a mean $f_o$ of 224.3 Hz decreasing to a mean $f_o$ of 201.1 Hz for women 60 years and over. Taken together, these results indicate that both age and sex are factors which must be controlled in describing fundamental frequency.
The variability of $f_o$ also has been reported in the literature. In the study by Horii (1975) the group standard deviation was 2.41 semitones. In the Hollien and Jackson study (1973), which used only young males, this variability was reduced slightly to a standard deviation of 1.6 tones. Stoicheff (1981) reported that standard deviation increased as a function of menopause. Her findings demonstrated that women’s voices premenopause and during menopause had a $f_o$ standard deviation of 3.92 and 3.97 semitones respectively, while post-menopausal women exhibited greater variability with a standard deviation of 4.48 semitones. Stoicheff interpreted these findings as evidence of decreasing phonatory control with advancing age. Mysak (1959) arrived at a similar conclusion for males based on a decreasing phonation-to-utterance time ratio which occurred with advancing age.

A more precise measure of fundamental frequency variability is the measurement of cycle-to-cycle perturbation of the glottal tone. When this variability is obtained from the prolonged phonation of a vowel at a constant pitch, the resultant measure is vocal jitter (Michel and Wnedahl, 1971). The formula for calculation of jitter (Horii, 1980) is:

$$\text{Mean jitter} = \frac{1}{(N-1)} X \sum_{i=1}^{N-1} |P_i - P_{i+1}|$$

where $N$ is the number of consecutive cycles analyzed and $P_i$ is the period of the $i^{th}$ cycle.

Coleman (1969) demonstrated that the amount of jitter present was in part a function of mean $f_o$. To counteract this effect, Horii (1980) recommended use of a jitter ratio which was the product of mean jitter divided by mean period. In the same study Horii reported a jitter ratio of .64 for normal young adults, which compared favorably with the values reported by Robbins, et al.
(1981) for normal male speakers (mean jitter ratio of .77). In the most recent study, Horii (1982) provided jitter ratio values for normal, young male adults for eight English vowels. The range of jitter ratio was .661 to .864 with a mean of .746.

Similar to \( f_0 \), vocal jitter also appears to be vulnerable to the effects of aging and of sex differences. Benjamin (1981) found greater variability in vocal jitter in older speakers (10.26) than in younger speakers (2.67) and in males (9.49) than in females (3.44).

A final frequency measure common in the analysis of vowels is formant frequency. This is the frequency at which energy transfer maximum occurs in the vocal tract (Fry, 1979). At least the first three formants (\( F_1 \), \( F_2 \), and \( F_3 \) respectively) are needed for accurate vowel identification (Foulkes, 1961). These three formants were identified first by Peterson and Barney (1952) for the vowels /i, ɪ, ɛ, æ, a, ə, u, ʊ, õ, and ɔ/, who also demonstrated that \( F_1 \), \( F_2 \), and \( F_3 \) were sensitive to age and sex differences, and that \( F_1 \) and \( F_2 \) were adequate for identification of non-retroflexed vowels. The vowel /i/ can serve as an example. The frequency of its \( F_1 \) for men was 270 Hz, for women 310 Hz and for children 370 Hz. Coleman (1971) demonstrated that this sex difference occurred independently of differences in \( f_0 \).

In frequency comparisons of esophageal or tracheoesophageal (T-E) phonation to normal voice, three generalizations may be made: first, alaryngeal speakers demonstrate greater variability of fundamental and formant frequencies. Second, regardless of sex or age, a marked reduction in fundamental frequency is a sequel to laryngectomy. Third, as mentioned earlier, formant values rise as a result of laryngectomy.
Angermeier and Weinberg (1981) investigated the control of fundamental frequency by esophageal speakers and normal speakers. Their results revealed that the standard deviation of $f_o$ for esophageal speakers was 1.5 to 8 times greater than that of the laryngeal group. In a similar study, Weinberg and Bennett (1972) found that both male and female esophageal speakers exhibited a standard deviation of fundamental frequency in the area of 4 semitones which was approximately twice the dispersion found in normal speakers. Robbins, et al. (1981) reported similar variability in $f_o$ but also included data for t-e speakers. Despite the overall conclusion that t-e speakers more closely resembled laryngeal speakers than esophageal speakers in voice production, these researchers found the t-e speakers had a much greater standard deviation (S.D.) of $f_o$ than even esophageal speakers: laryngeal speakers S.D.- 12.69, esophageal speakers S.D.- 18.17, and t-e speakers S.D.- 54.64 semitones. In a second study of alaryngeal voice supported by pulmonary air, Robbins, et al. (1982) found that two subjects who had undergone the Staffieri procedure demonstrated vocal $f_o$ standard deviations of 2.25 and 6.09 semitones. Vocal jitter as a measure of the stability of the fundamental frequency also revealed the poorer control over $f_o$ exerted by both esophageal and p.a.a. speakers. To date only Robbins, et al. (1981) have addressed this issue. They found a very wide difference between normal and t-e speakers and between t-e speakers and esophageal speakers: normal- mean jitter ratio of 7.74 with a S.D. of 5.13; t-e- mean jitter ratio of 51.35 with a S.D. of 46.74; and esophageal- mean jitter ratio of 182.45 with a S.D. of 97.54. These data were only one of fifteen variables entered into a linear discriminant analysis in order to identify by which variables laryngectomy, t-e, and esophageal speakers could be differentiated. Of these fifteen variables,
vocal jitter was fifth in potency of discrimination and highest of the frequency related dependent variables.

Apart from the control of $f_0$, another common characteristic of alaryngeal voice production is a lowering of $f_0$ for both males and females. The male esophageal $f_0$ is approximately one half that of male laryngeal speakers, or in the vicinity of 65 Hz (Hartman, 1979; Hyman, 1979; Christensen and Weinberg, 1976; Curry and Snidcor, 1961). The female esophageal voice exhibits a similar decline to approximately 87 Hz (Weinberg and Bennett, 1972), which, while higher than the male esophageal voice, is only one third the normal female vocal $f_0$. This lower $f_0$ of esophageal speakers generally has been attributed to the relatively greater flaccidity and weight of the vibratory site, the pharyngo-esophageal (p-e) segment, as compared to the vocal folds. The fact that this site represents a surgical remnant which may differ greatly across individuals also has served to explain the greater variability in control of the fundamental of the alaryngeal population as opposed to normal, laryngeal speakers (Weinberg, 1980a).

While this phenomenon of lowered $f_0$ has appeared consistently, the inter-subject variability has appeared just as consistently, and has been interpreted by some researchers as indicative that the mechanism for production and control of $f_0$ is qualitatively different than that of the normal voice (Smith, et al., 1978). This interpretation gains credence from the accumulating, but sparse data concerning paa vocal $f_0$. In one study Robbins, et al. (1982) reported a mean $f_0$ of 48.4 Hz for speakers using Staffieri’s technique, placing the subjects for this study well below even esophageal speakers. In assessing the perceptual correlate of $f_0$, pitch, however, Graner, et al. (1982) found that listeners reported only moderately depressed pitch for speakers
using the Staffieri procedure. In Robbins, et al. (1981) the mean $f_o$ for the paa speakers (all Blom-Singer prosthesis users) was only 1 Hz lower than the mean $f_o$ for the normal group (102 Hz and 103 Hz), but, as cited above, the mean jitter ratio was much greater for the t-e subjects.

In general the paa speakers appear to produce a somewhat higher $f_o$ than the esophageal speakers and better control of $f_o$; despite the use of the p-e segment as the vibratory site for both populations. A possible explanation of this occurrence is that the pressure exerted by intraesophageal air on the p-e segment may serve to tense the segment and, therefore, to elevate $f_o$. The coupling of the respiratory system with the esophagus through the surgical shunt may allow greater preservation, or even an increase of, intraesophageal air pressure (Weinberg, et al., 1982) by which the p-e segment is tensed for greater periods of time.

To date the information concerning production and control of the fundamental by paa speakers is limited to male subjects. The vocal performance of female paa speakers remains a matter of future research.

In terms of resonant characteristics, the available data for both female and male paa speakers is limited. The Staffieri patients reported by Robbins, et al. (1982) represent only two male subjects and constitute the sole reference to formant values currently available in the literature. In general, while the upward formant shift common to esophageal voice compared to normal voice did occur, the Staffieri speakers did not demonstrate as dramatic a shift as was evident for esophageal speakers.
TABLE 2-1. Representative formant l-3 values for normal, esophageal, and tracheoesophageal male speakers.

<table>
<thead>
<tr>
<th></th>
<th>Normal(^1)</th>
<th>Esophageal(^2)</th>
<th>T-E(^3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>/i/</td>
<td>F(_1) 270</td>
<td>F(_2) 401</td>
<td>F(_3) 312</td>
</tr>
<tr>
<td></td>
<td>F(_2) 2290</td>
<td>F(_2) 2684</td>
<td>F(_3) 1920</td>
</tr>
<tr>
<td></td>
<td>F(_3) 3010</td>
<td>F(_3) 3067</td>
<td>F(_3) 2533</td>
</tr>
<tr>
<td>/a/</td>
<td>F(_1) 730</td>
<td>F(_2) 984</td>
<td>F(_3) 706</td>
</tr>
<tr>
<td></td>
<td>F(_2) 1090</td>
<td>F(_2) 1357</td>
<td>F(_3) 1508</td>
</tr>
<tr>
<td></td>
<td>F(_3) 2440</td>
<td>F(_3) 2830</td>
<td>F(_3) 2334</td>
</tr>
<tr>
<td>/u/</td>
<td>F(_1) 300</td>
<td>F(_2) 459</td>
<td>F(_3) 387</td>
</tr>
<tr>
<td></td>
<td>F(_2) 870</td>
<td>F(_2) 1213</td>
<td>F(_3) 992</td>
</tr>
<tr>
<td></td>
<td>F(_3) 2240</td>
<td>F(_3) 2666</td>
<td>F(_3) 2376</td>
</tr>
</tbody>
</table>

\(^1\) Peterson and Barney (1952)
\(^2\) Sisty and Weinberg (1972)
\(^3\) Robbins, et al. (1982)

The conventional explanation for the formant shift was posited first by Sisty and Weinberg (1972). They theorized that the laryngectomy and subsequent use of the p-e segment for phonation produced a shorter vocal tract which would result in higher resonant frequencies. Christensen and Weinberg (1976) suggested that total laryngectomy also affected the articulatory gestures necessary for intelligible speech. Since articulation also affects the resonant characteristics of the vocal tract, and since many proficient esophageal
geal speakers utilize plosive injection (the use of the high intraoral pressure associated with stop consonants to charge the esophagus) throughout their speech (Diedrich, 1968; Moolensar-Bijl, 1953), the upward shift in formant frequencies may be the result of the combination of a shortened vocal tract and of modified articulatory gestures (Tikovsky, 1965). This would help resolve why the speakers reported by Robbins, et al. (1982) failed to exhibit the expected degree of formant shift: the need for articulatory gestures to charge the esophagus no longer existed; therefore, as articulation became more normal the only factor acting to alter resonatory characteristics was the shortened vocal tract.

Just as the average values for formants 1-3 are higher for females than male laryngeal speakers, so also are the first three formants for female esophageal speakers higher than they are for their male counterparts, as well as higher than those of female laryngeal speakers. The upward formant shift in the female esophageal speaker is not as great proportionately as the shift found in the male esophageal group. It can be hoped that since male t-e speakers appear to experience less formant elevation that male esophageal speakers, the female t-e speaker likewise would experience a smaller elevation. This would produce formant patterns more closely resembling normal patterns. As yet, however, no data on female t-e speakers' formant structure has been published.
### TABLE 2-2. Representative formant 1-3 values for normal and esophageal female speakers.

<table>
<thead>
<tr>
<th></th>
<th>Normal(^1)</th>
<th>Esophageal(^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>/i/</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(F_1)</td>
<td>310</td>
<td>390</td>
</tr>
<tr>
<td>(F_2)</td>
<td>2790</td>
<td>2925</td>
</tr>
<tr>
<td>(F_3)</td>
<td>3310</td>
<td>3627</td>
</tr>
<tr>
<td><strong>/a/</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(F_1)</td>
<td>850</td>
<td>1031</td>
</tr>
<tr>
<td>(F_2)</td>
<td>1220</td>
<td>1432</td>
</tr>
<tr>
<td>(F_3)</td>
<td>2810</td>
<td>3012</td>
</tr>
<tr>
<td><strong>/u/</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(F_1)</td>
<td>370</td>
<td>435</td>
</tr>
<tr>
<td>(F_2)</td>
<td>950</td>
<td>1134</td>
</tr>
<tr>
<td>(F_3)</td>
<td>2670</td>
<td>2835</td>
</tr>
</tbody>
</table>

1 Peterson and Barney, 1952
2 Sisty and Weinberg, 1972
**Intensity Related Measures**

Just as substantial changes in the frequency domain occur following laryngectomy, similar changes also occur in the intensity domain. These changes are documented by reporting mean or median vocal intensity and its standard deviation in conversation or reading and mean vocal shimmer and its standard deviation. The dominant change is a reduction in intensity attributable directly to the absence of pulmonary air to support voice (Diedrich, 1968; Diedrich and Youngstrom, 1966; Blom and Singer, 1979). The intensity of voice is largely a function of rate of air flow through the glottis or neoglottis. For the normal, laryngeal speaker, this rate of flow arises from a controlled release of expiratory reserve and tidal air. Herein the laryngectomee deviates from the normal subject. First, expiratory reserve, which is the amount of air a person can exhale over and above tidal air, averages about 1500 cc (Zemlin, 1968). While the esophageal speaker may possess the same expiratory reserve as his laryngeal counterpart, this pulmonary air no longer subserves speech. Instead the esophageal speaker must rely on the capacity of the esophagus for storing air. Esophageal air capacity is approximately 70 cc, of which probably only 25 cc support voice (Diedrich and Youngstrom, 1966). This results in a 95% reduction in air available to support voice. Second, respiration, which is the physiological basis for voice, is subject to precise muscular control via the diaphragm, internal intercostals and abdominal muscles (Zemlin, 1968; Lieberman, 1977). Ejection of air from the esophagus appears to rely on the inherent elasticity of the esophagus and, therefore, is subject to less precise control (Diedrich and Youngstrom, 1966). Third, the larynx itself serves primarily as a valve segregating the upper and lower respiratory tracts. As such the larynx acts intrinsically in the control of air flow both from and into the
lungs. In occluding the airway during phonation, the vocal folds interfere with the flow of air much less than does the p-e segment. Specifically the impedance offered by the p-e segment appears to be much greater than that offered by the laryngeal mechanism (Weinberg, 1980a; Weinberg, et al., 1982). Together these three factors suggest that alaryngeal voice should be substantially less intense than normal voice and the accumulated data support such a conclusion. The support, however, must be interpreted in light of the type of task used to measure intensity, the age and sex of the subjects, and the mouth to microphone distance.

Ptacek, et al. (1966) asked four groups of subjects to produce the vowel /a/ with maximum intensity. The four groups were divided by age and sex: young males with a mean age of 27.6 years; young females with a mean age of 23.5 years; old males with a mean age of 76.9 years; and old females with a mean age of 76.9 years. Intensity was measured 12 inches (30 cm) from the source and averaged over three trials. The two younger groups were not significantly different from each other (males 105.8 dB SPL, females 106.2 dB SPL) but were significantly more intense than the two older groups which were significantly different from each other as well (males 100.5 dB SPL, females 98.6 dB SPL). While this study did not address possible differences between normal and alaryngeal voice, it did emphasize the need to control for both age and sex in assessing vocal intensity.

Traditionally, the main recourse of investigators in controlling sex differences in the alaryngeal population has been to eliminate female subjects from consideration. While the reported incidence of laryngeal cancer in women varies with the investigation, only one sixth (American Cancer Society, 1975) to one tenth (Weinberg and Bennett, 1972) of the laryngeal-
tomoe population are women. Since the total number of laryngectomies performed annually in the United States ranges from 9000 to 11000 (Singer and Blom, 1980), this translates to less than 1900 performed on women yearly. Stated briefly, researchers have failed to gather data on the intensity characteristics of female alaryngeal voice because a representative sample is very difficult to obtain.

While the intensity data for female esophageal speakers are absent, the data for male esophageal speakers are remarkably consistent. Overall the intensity of esophageal voice is recognized as 6 to 10 dB less than the intensity of normal voice and the range of intensity for the esophageal population is approximately half that of the normal population (Hartman, 1978; Hyman, 1978; and Martin, 1978).

The limited, available information regarding intensity characteristics of paa voice is also limited to male speakers and indicates that the maximum intensity of paa voice approximates that of normal voice (Robbins, et al., 1981; Robbins, et al., 1982). Table 2-3 summarizes these findings.

The control of vocal intensity appears to be very different across the three groups of speakers. Robbins, et al. (1981) presented data for the standard deviation (S.D.) of oral reading intensity as well as for vocal shimmer and its standard deviation (Table 2-4). Shimmer is the cycle-to-cycle variability in intensity exhibited during prolongation of a vowel at constant pitch and loudness (Michel and Wendahl, 1971). The formula to derive this measure is (Horii, 1980):

\[
\text{Mean Shimmer in dB} = \frac{20}{N-1} \sum_{i=1}^{N-1} \log_{10} \frac{A_i}{A_{i+1}}
\]

where \( A_i \) is the peak amplitude of the \( i^{th} \) cycle and \( N \) is the number of continuous cycles analyzed.
TABLE 2-3. Intensity characteristics of esophageal, t-e and normal male speakers.

<table>
<thead>
<tr>
<th></th>
<th>Esophageal</th>
<th>T-E</th>
<th>Normal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Intensity</td>
<td>62.4 dB SPL(^1)</td>
<td>79.4 dB A(^3)</td>
<td>69.3 dB A(^3)</td>
</tr>
<tr>
<td>Range</td>
<td>20 dB(^2)</td>
<td>17 dB(^4)</td>
<td>45 dB(^2)</td>
</tr>
</tbody>
</table>

\(^1\) Hoops and Noll, 1969  
\(^2\) Hartman, 1978  
\(^3\) Robbins, et al., 1981  
\(^4\) Robbins, et al., 1982, (mean to peak amplitude)

TABLE 2-4. Intensity control of esophageal, t-e, and normal male speakers as demonstrated in intensity standard deviation, vocal shimmer and shimmer standard deviation.

<table>
<thead>
<tr>
<th></th>
<th>Esophageal</th>
<th>T-E</th>
<th>Normal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intensity S.D.</td>
<td>4.8</td>
<td>2.13</td>
<td>2.94</td>
</tr>
<tr>
<td>Mean Shimmer</td>
<td>27.15</td>
<td>10.55</td>
<td>4.29</td>
</tr>
<tr>
<td>Shimmer S.D.</td>
<td>10.51</td>
<td>6.79</td>
<td>1.71</td>
</tr>
</tbody>
</table>

\(^1\) Robbins, et al., 1981
The Robbins' data indicate that, while in terms of intensity S.D. the laryngeal and paa speakers are very similar, the cycle-to-cycle variability of the paa speaker is much greater. This finding is not totally unexpected given the quasiperiodic nature of p-e segment vibration. It is interesting, however, that esophageal speakers who utilize the same vibratory source exhibit nearly three times more cycle-to-cycle variability. The most obvious explanation for the discrepancy among the three groups is that variability of intensity is not solely dependent on the energy source (i.e. air supply), but on the vibratory site as well. The esophageal speaker lacking both the laryngeal vibratory mechanism and respiratory support for phonation exhibits the greatest variability and reduction in intensity. The t-e speaker can utilize respiratory air but uses it in a more massive and resistive vibratory body (Weinberg, 1980a; Weinberg, et al., 1982); and, therefore exhibits variability that is midway between esophageal and normal voice, but without the reduction of intensity exhibited by esophageal speakers.

This data in intensity characteristics of paa voice must be interpreted with caution as it rests on only three interconnected articles or papers authored by various combinations of the following people: E. Blom, H. Fisher, J. Hillenbrand, Y. Horii, J. Logemann, J. Robbins, M. Singer, and B. Weinberg. While Robbins, et al. (1981) reported a t-e sample size of 15, Robbins, et al. (1982) reported a sample size of 2, and Weinberg, et al. (1982) reported a sample of 5 t-e speakers. This would constitute a total sample of only 22 speakers if no overlap in sampling could be assumed. Unfortunately such an assumption is tenuous. Robbins, et al. (1981) reported the largest single group of subjects and noted "All t-e subjects had undergone TEP by Dr. Mark Singer" (p.1). While the investigating techniques of the above authors are
refined and sophisticated, the consanguinity of the subjects and of the researchers indicates a need for independent effort to provide additional information on the intensity characteristics of voice restored through puncture techniques.

Temporal Measures

The effects of total laryngectomy on the temporal aspects of voice largely arise from the same source as do intensity effects, to wit, removal of pulmonary vocal support. In general the esophageal speaker demonstrates a considerable decrement in almost every measure used to gauge duration of voice. The exception to this generality is that esophageal speakers, who have lost their "voice boxes", produce proportionately more voicing in their speech than do normal speakers. Christensen and Weinberg (1976) and Christensen, et al. (1978) demonstrated that vowel duration time, hence voiced time, was proportionately greater in esophageal speakers than in normal speakers. The apparent cause of this phenomena was that esophageal speakers exhibited a significantly shorter voice onset time for vowels in a voiceless consonant environment. The authors concluded that alaryngeal speakers possessed poorer temporal control of voice than did laryngeal speakers.

A similar conclusion based on different tasks had been posited by Berlin in 1963. In what has come to be considered a landmark study (Weinberg, 1980a), Berlin measured four different vocal behaviors of esophageal speakers and related his measures to patient progress in attaining functional esophageal voice. Berlin's four measures (consistency of phonation, latency of phonation, duration of phonation, and number of syllables per air intake) were temporal in nature. Not surprisingly, Berlin found that individuals who could produce voice consistently on demand held a better prognosis for achieving proficient
esophageal voice than those individuals who could not do so. Berlin's latency task was measured as the time interval between the cue to phonate and the production of audible phonation. Phonation to support conversational speech should not require a lengthy nonphonatory period to precede voice production. Berlin found that proficient esophageal speakers could reduce prephonatory latency to less than 0.5 seconds. For measuring the duration of phonation, Berlin required his subjects to produce the vowel /a/ for as long as possible on only one air intake. Overall, he found that a mean duration of 1.8 seconds or greater correlated highly with the acquisition of functional voice. In his group of proficient speakers, Berlin reported a range of vowel duration of 1.8 to 4.0 seconds with a mean of 2.8 seconds and a standard deviation of .52 seconds. In syllable per air intake, Berlin's subjects produced the syllable /da/ as often as possible before expending available air. The proficient speakers' mean performance in this task was 8.6 syllables with a standard deviation of 2.24 syllables and a range of 4 - 14 syllables.

Despite the absence of normal laryngeal speakers from Berlin's sample, the nature of Berlin's tasks does much to reveal the limitations of esophageal voice. Berlin's first two tasks, consistency and latency of phonation, essentially have no counterparts in normal voice research, primarily because they are not issues in normal voice production. The normal laryngeal speaker produces voice when desired and without appreciable delay because voice production is a function overlaid on a basic physiological process, respiration. Since such is not the case for the alaryngeal speaker who somehow must inflate the esophagus, failure to achieve rapid, efficient esophageal inflation can yield measurable delays in, or failure of, phonation. Berlin's final two measures deal with the difference between pulmonary vital capacity and the
air reservoir capacity of the esophagus (an esophageal speaker must replace a 1500 cc air source with a 70 cc air source). The vowel prolongation task may be the most telling in this instance because a direct comparison can be made with the performance of normal speakers and t-e speakers. Ptacek and Sander (1963) asked two groups of adults to prolong /a/ for as long as possible on one breath. The groups were divided on the basis of sex: 40 males with a median age of 23 years and 40 females with a median age of 20.9 years. These researchers found sex differences in that males could phonate substantially longer than females; but of greater pertinency in an esophageal to normal comparison, even the mean duration time of the females (17 seconds) far surpassed Berlin's all male alaryngeal sample's performance (2.8 seconds). Robbins, et al. (1981) again provide a comparison of esophageal, t-e, and normal speakers (all male). The mean duration of /a/ for these three groups respectively was 1.92 seconds, 12.16 seconds, and 21.83 seconds. Very clearly the paa speakers surpassed the capabilities of the esophageal speakers but also fell well short of the level of the normal speakers.

To date the literature has provided no data on syllable repetition for normal or paa speakers and, therefore, comparisons across groups are not possible. Even if data were available, the interpretation of comparisons would be difficult. In the /da/ repetition task Berlin observed that the phonatory time involved in this measure often exceeded the phonatory time obtained during vowel prolongation. Berlin resolved this incongruency by hypothesizing that some esophageal speakers could utilize the high intraoral air pressure associated with plosive consonants to replenish esophageal air. Such a conclusion had been reached by assorted other investigators (Moolensar-Bijl, 1953; Weinberg and Bosma, 1970; and Damste, 1979) and the act of plosive injection
was measured by Snidecor and Isshiki (1965). A similar behavior has never been recorded in the speech of normal or paa subjects who possess adequate pulmonary air support for speech and do not require such a compensatory mechanism.

Berlin's data support the conclusion that esophageal speakers are less able to sustain phonation than are paa or normal speakers. Based on these data one could predict that esophageal speakers would exhibit greater difficulty in maintaining voice. Two related measures, rate of oral reading and the ratio of pause time to total time in oral reading, lend credence to this prediction. Rate of oral reading, usually expressed as words per minute, and the pause time to total time ratio, usually expressed as a percentage, reflect the efficiency with which a speaker can support phonation necessary for speech. Difficulty in phonation yields decreased number of words per minute and an increase in total pause time. The normal, adult speaker reads aloud at a rate of approximately 166 words per minute (Fairbanks, 1960) with a ratio of pause time to total time of 14% (Weinberg, et al., 1980b) to 17.73% (Robbins, et al., 1981). Table 2-5 summarizes the information available for esophageal and paa speakers. Apparent from this table is the considerable variation in performance not only across groups, but also within groups, for rate of oral reading and percentage of pause time. One possible factor contributing to this variability is the level of speaking proficiency of the subjects. Snidecor and Isshiki (1965), Hoops and Noll (1969), and Weinberg, et al. (1980b) all applied some form of screening mechanism for speech proficiency in subject selection while Robbins, et al. (1981) and Graner, et al. (1982) did not. As evidenced in Table 2-5, the performance of subjects from studies with no level of speaker
adequacy used was generally inferior to the performance of subjects from studies with a proficiency criterion established.

Nevertheless, viewed in toto, these results again reveal that both esophageal and paa speakers exhibit reductions in vocal abilities when compared with normal speakers and that paa speakers are superior to esophageal speakers.

**Perceptual Measures**

To this point the literature reviewed for this investigation has described events associated with the act of producing voice or with the disturbance of the acoustic medium necessary for the transmission of voice. These events are precursors to the goal of voice production, i.e., the exchange of information from one individual to another. The following review concerns the receptive end of the speech chain; the effect vocal communication has upon the listener.

The foregoing material has described the effects of total laryngectomy with esophageal or paa voice restoration upon the production of the speech signal, but has not addressed how the changes in voice affect the listener. In general, two perceptual phenomena are affected, intelligibility and the perception of vocal quality. As might be anticipated from the altered physical structure of alaryngeal voice, the intelligibility and the vocal quality of alaryngeal speakers consistently have been found to be inferior to that of normal, laryngeal speakers. This section will review the degree of decrement affecting alaryngeal intelligibility and vocal acceptability and identify those factors commonly reported as contributing to these decrements.

**INTELLIGIBILITY:** This perceptual measure is normally expressed as a percentage correct and is assessed by recording a speaker reading lists of individual words or sentences and playing that recording to a group of listeners who
TABLE 2-5. Comparison of mean reading rate and mean percent of pause time in reading for esophageal and paa speakers.

<table>
<thead>
<tr>
<th>Speaker type</th>
<th>Rate of oral reading</th>
<th>% of pause time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Robbins, et al. (1981)</td>
<td>99.08</td>
<td>36.06</td>
</tr>
<tr>
<td>paa</td>
<td>127.48</td>
<td>24.20</td>
</tr>
<tr>
<td>Weinberg, et al. (1980)</td>
<td>______</td>
<td>23.00</td>
</tr>
<tr>
<td>Hoops and Noll (1969)</td>
<td>114.3</td>
<td>______</td>
</tr>
<tr>
<td>Snidecor and Isshiki (1965)</td>
<td>116.83</td>
<td>______</td>
</tr>
<tr>
<td>Robbins, et al. (1982)</td>
<td>125.00</td>
<td>17.45</td>
</tr>
<tr>
<td>paa</td>
<td>77.35</td>
<td>______</td>
</tr>
</tbody>
</table>
identify what they hear. The percentage reflects the degree of agreement between the listeners and the original list.

Table 2-6 summarizes the results of approximately 20 years research into alaryngeal intelligibility. The initial three studies show strong agreement among themselves, but contrast sharply with the latter two studies. A superficial explanation for this difference might be that esophageal speakers have improved over the preceding two decades. A more probable explanation emerges from the subject selection criteria of each study. In the initial three articles any esophageal speaker who could complete the respective intelligibility task was utilized. Horii and Weinberg (1975), and Clark and Stemple (1982), however, applied stricter standards in subject selection. Clark and Stemple selected their alaryngeal speakers by providing taped samples of prospective speakers to a panel of three speech pathologists who identified those samples which represented above average alaryngeal voice production. Horii and Weinberg selected their subjects based on automaticity of speech, absence of extraneous noise, and pleasant vocal quality. They concluded that the speech of their subjects was "among the best alaryngeal speech the authors had ever heard." (p.414) The apparent cause for the considerable difference in reported intelligibility, then, lies in the proficiency of the speakers. Above average alaryngeal speakers approximate the intelligibility of normal laryngeal speakers. The performance of average esophageal speakers obviously fell well below the performance of their normal counterparts.
TABLE 2-6. Summary of intelligibility scores for normal, esophageal and T-E speakers in quiet speaking and listening conditions.

<table>
<thead>
<tr>
<th>Study</th>
<th>Normal</th>
<th>Esophageal</th>
<th>T-E</th>
</tr>
</thead>
<tbody>
<tr>
<td>McCroskey and Mulligan(1963)</td>
<td>58.2%</td>
<td>98.2%</td>
<td></td>
</tr>
<tr>
<td>Tikovsky (1965)</td>
<td>92.7%</td>
<td>61.6%</td>
<td></td>
</tr>
<tr>
<td>Creech (1966)</td>
<td></td>
<td>62.6%</td>
<td></td>
</tr>
<tr>
<td>Horii and Weinberg (1975)</td>
<td>98%</td>
<td>98%</td>
<td></td>
</tr>
<tr>
<td>Clark and Stemple (1982)</td>
<td>100%</td>
<td>97.5%</td>
<td>99.0%</td>
</tr>
</tbody>
</table>

1 Vowel intelligibility score

This conclusion concerning the proficiency of alaryngeal speakers holds with one important proviso; the scores reported in Table 2-6 were obtained in ideal listening conditions (i.e., in the absence of appreciable background noise). With the addition of noise in the listener's environment, the intelligibility scores of all speakers, normal and alaryngeal, deteriorate; however, the scores of alaryngeal speakers deteriorate more rapidly. This deterioration may be attributed largely to the decreasing intelligibility of consonants in noise. Horii, et al. (1971) described the decrease of intelligibility of the speech of normal speakers in the presence of increasing noise as a function of two curves, a curve of consonant intelligibility and a curve of vowel intelligibility. For normal speakers the slope of the vowel curve was 4% per decibel of noise, while the slope of the consonant curve was 2.5% per dB of noise. The Horii and Weinberg study of 1975 is essentially a duplication of Horii, et al. (1971) using esophageal speakers. The curve of vowel intelligibility decline in noise for esophageal speakers was identical to the curve for normal speakers, 4% per dB. The curve of consonant intelligibility decline in noise, however, was twice as steep as that of normal speakers, 5% per dB. These results indicate that as
background noise increases, the alaryngeal speaker's intelligibility deteriorates more rapidly than does the normal speaker's intelligibility. The Clark and Stemple study (1982) provided additional evidence for this phenomena. When the signal to noise ratio in that study was increased from 0 to -5 dB, intelligibility fell to 94.5% for normal speakers, 84.5% for esophageal speakers and 75.5% for t-e speakers. The data strongly support an overall interpretation that due to the altered physical characteristics of alaryngeal voice, the addition of competing acoustic signals markedly reduce a listener's ability to interpret the content of esophageal and paa voice.

Two further questions arise from the data: 1) What is the nature of the intelligibility reduction? and 2) Why does paa intelligibility deteriorate more rapidly that that of esophageal voice? Sacco, et al. (1967) provided insight into the nature of errors producing the intelligibility reduction. In their work, 19 esophageal speakers produced 256 different consonant-vowel syllables constructed from 16 consonants and the vowel /a/. These productions were assessed for intelligibility by 10 judges and the results displayed in a confusion matrix. This analysis revealed that over 50% of the errors were due to voicing confusion, and that unvoiced consonants (consonants not characterized by vocal fold vibration) were twice as likely to be misidentified as their voiced counterpart. This study indicates that esophageal speakers possess poorer control over voice onset than do normal speakers and that the poorer control is exhibited as the early initiation of voicing in the context of unvoiced consonants. This phenomena was identified and measured by Christensen and Weinberg (1976), as cited. The addition of noise degrades the already reduced acoustic cues, differentiating voiced and voiceless consonants and, therefore, exacerbates the listener's problem of interpreting an abnormal vocal signal.
There remains some question as to why the t-e speaker in the Clark and Stemple study (1982) demonstrated a more accelerated decline in intelligibility than did the esophageal speakers when the data cited earlier indicate that the t-e subject should have produced a more nearly normal voice. The lack of available data only allows for speculation on this subject. The Clark and Stemple article is the only article to date which has addressed the topic of intelligibility of paa voice. The total number of paa speakers in that study's sample is one. That subject's speech was recorded in silence, then noise was added to produce the described signal to noise ratios. Due to the limited sample size, the generalizability of Clark and Stemple's findings to the population of paa speakers would be reckless. Additionally, the Robbins study of 1981 indicated that t-e speakers have a wider range of intensity and better control over intensity that do esophageal speakers. Both paa speakers and esophageal speakers produce extraneous noise during the production of speech. One major source of this noise is the exhalation of air through the tracheostoma (Hyman, 1979). It is reasonable to assume that when speaking in a quiet setting, both types of speakers were producing stoma noise in addition to speech and that the paa speaker had a considerable amount of the total range of intensity unused because a quiet setting did not require intense speech. For both speakers, the act of respiration created noise in the environment; but for the t-e speaker, whose voice would be closer to its minimum intensity in such conditions, the stoma noise would produce a smaller signal to noise ratio than would be the case for the esophageal speaker whose voice would be closer to its maximum attainable intensity. The addition of noise to the tape recording, therefore, would degrade both voices equally, but maintain the imbalance already created by stoma noise. A more realistic approach to the matter of
paa intelligibility would be to record speakers in environments varying in the amount of noise present in order to determine the degree to which speakers may adapt to preserve intelligibility. This, however, is the stuff of future research.

ACCEPTABILITY: Definitive conclusions about the factors contributing to the acceptability of alaryngeal voice are illusive in the literature. The major areas of agreement are that normal voice is more acceptable than any form of alaryngeal voice (Clark and Stemple, 1982; Green and Hults, 1982; Rusnov, et al., 1981; and Bennett and Weinberg, 1973), that paa voice is more acceptable than esophageal voice or electrolaryngeal voice (Clark and Stemple, 1982), and that esophageal voice is more acceptable than electrolaryngeal voice (Green and Hults, 1982, Bennett and Weinberg, 1973), at least when the judges of acceptability are speech pathologists (McCroskey and Mulligan, 1963). A hierarchy of acceptability is plain. The hierarchy descends from normal voice to paa voice, to esophageal voice, to electrolaryngeal voice. What is not plain are those elements of the voice signal which determine the hierarchy. Studies attempting to describe those factors active in influencing vocal acceptability generally identify the same set of variables, but not in the same order of importance, probably because of variation in the measurement of acceptability and of the contributing factors. Clark and Stemple (1982), Hoops and Noll (1969) and Shipp (1967) employed ordinal scales to determine acceptability, while Bennett and Weinberg (1973) constructed a seven-point equal appearing interval scale for that task. Clark and Stemple related acceptability to only one other factor, intelligibility, and in the end could not resolve the paradox that t-e speech was consistently judged superior to the two other forms of alaryngeal voice in acceptability, but was inferior in intelligibility in two of
three listening conditions. Hoops and Noll used 7 dependent variables (mean and standard deviation of the vocal fundamental frequency and of vocal intensity, period-to-period perturbation of the vocal fundamental, mean words per minute, and mean words per sentence per minute). After completing a multiple correlation these authors found that mean words per minute and mean words per sentence per minute accounted for more than 96% of the variance in vocal acceptability, relegating mean \( f_o \) and mean vocal intensity to secondary roles.

In a very similar study, however, Shipp (1967) had found that mean \( f_o \) was the most potent predictor of vocal acceptability, followed by mean words per minute, and the proportion of periodic sound present in an utterance. Bennett and Weinberg, who went to greatest pains to quantify acceptability, simply asked their subjects to report what factors influenced their judgements. In differentiating esophageal voice from normal voice, the most common factor cited by the judges was the voice "quality did not sound normal" (p.612), followed by the perception of slow speaking rate and low pitch.

**Summary**

It is hoped that the final outcome of the present investigation will shed some light on the vocal characteristics by which alaryngeal speakers differ among themselves and from normal, laryngeal speakers and how these factors affect the listener's perception of voice. This chapter has delineated four areas of voice research (the frequency domain, the intensity domain, the temporal domain, and the perceptual domain) and attempted to identify pertinent measure in each of these areas. The following is a listing of the variables in each domain selected for inclusion in the present study:
Frequency:
Fundamental Frequency of /a/
Range of $f_0$ among quantal vowels /i,a,u/
$F_1$ and $F_2$ peak frequencies for quantal vowels /i,a,u/
$F_2:F_1$ ratios for the quantal vowels /i,a,u/

Intensity:
Maximum intensity of comfortable /a/ production
Range of intensity between least and most intense /a/

Temporal:
Latency of phonation
Consistency of phonation
Duration of phonation
Mean number of syllables produced on one air intake
Words per minute in oral reading
Ratio of pause time to total time in oral reading

Perceptual:
Vocal acceptability measured on an equal-appearing interval scale
CHAPTER III

Procedures

The Problem

In overview, the procedures of this study were applied in order to obtain representative voice samples from three groups of speakers: esophageal, pulmonary-assisted-alaryngeal (paa) and normal laryngeal speakers. These voice samples then were assessed along a number of physical measures and one perceptual measure. Chapter III has been divided into five sections: 1) subjects contributing voice samples; 2) recording instrumentation; 3) equipment and methodology for the measurement of the physical characteristics of voice; 4) equal-appearing interval scale — construction and application; and 5) chapter summary.

Speaking Subjects

A total of 45 subjects supplied voice samples for this study. They may be divided into three groups: normal laryngeal speakers, paa speakers, and esophageal speakers. Each group contained 15 individuals, 10 males and 5 females. The speaking adequacy of each speaker was judged by a panel of three speech pathologists who worked routinely with alaryngeal speakers. The panel was comprised of a speech pathologist at Riverside Methodist Hospital, Columbus, Ohio, the Director of Speech Pathology, Grant Hospital, Columbus, Ohio, and this investigator. The ordinal scale of proficiency used by the panel (see
Appendix A) was assigned numerical values: 0 - Below average, 1 - Average, 2 - Above average, and 3 - Superior.

NORMAL LARYNGEAL SPEAKERS: Speakers in this group were judged by the panel of speech pathologists as demonstrating normal voice for their age and sex. The overall age of this group was 58.27 years. The mean age of male speakers was 58.7 years with a range of 46 years to 72 years. The mean age of the female speakers was 57.4 years with a range of 50 to 62 years.

PAA SPEAKERS: The speakers in this group were judged by the panel as proficient users of paa voice. The poorest speakers in this group was judged as Average by two judges and Above Average by the remaining judge. The two best speakers were judged as Superior by all judges. The median proficiency level for all paa speakers was 2.75 in a range of one to three. For the male paa speakers the median proficiency level was 2.17 in a range of one to three. For the female paa speakers, the median was 2.93 in a range of one to three.

The overall mean age of the paa group was 59.8 years. For the men, the mean age was 60.3 years with a range of 37 to 82 years. For the women, the mean age was 58.8 years with a range of 56 to 63 years. On the average, the paa speakers had been fitted with a voice prosthesis for 16.45 months. The speaker who had the least experience with paa voice was a woman who had received her prosthesis three weeks prior to recording. The speaker with the most experience was a man who had received his prosthesis 36 months prior to the recording. Thirteen paa speakers used the Blom Singer Voice Prosthesis (BSVP), and two men used Shapiro's modification of the BSVP (SP).

ESOPHAGEAL SPEAKERS: The speakers in this group were judged to be proficient users of esophageal voice. The poorest speaker in this group was judged as Average by all judges. The best speaker was judged as Superior by all
judges. The median proficiency level for all esophageal speakers was 2.5 in a range of one to three. For the male esophageal speakers the median proficiency level was 2.5 in a range of one to three, and for females the median was 2.6 in a range of one to three.

The mean age of the esophageal group was 57.5 years. For the male speakers, the mean age was 58.7 years with a range of 42 to 70 years. For the female speakers the mean age was 55 years with a range of 24 to 73 years. The average esophageal speaker had undergone total laryngectomy 95.5 months prior to recording, while the most recent laryngectomy occurred five months prior to recording and the laryngectomy of longest standing was 22 years, 2 months prior to recording.

**Recording Instrumentation**

The maintainence of a known reference signal was crucial to the accurate measurement of the intensity parameters of voice; therefore, the following procedures were instituted to insure that "0" V.U. remained equivalent to 84.5 dBSPL. For all the subjects except the six recorded at Wichita State University and the three recorded in Newark, New Jersey, the same recording and calibration equipment was used: a TEAC A-350 stereo cassette deck, TDK SA C 60 Super Alivyn cassettes, a Realistic pencil-style electret condenser microphone (frequency response range of 20-1600 Hz with a sensitivity of 74+3dB at 1 KHz) mounted on a custom built headset, a Hewlett Packard Audio Oscillator, Model 200 AB, and a Hewlett Packard 350 D Attenuator Set. Because esophageal, paa and normal subjects contributed to the sample, a wide range of intensity across subjects was expected. A consistent finding in past alaryngeal voice research was that esophageal speakers were 6-10 dB less intense than normal speakers (Weinberg, et al., 1980b) and the mean intensity of paa
speakers tended to approximate normal intensity (Robbins et al., 1981). This range of intensity presented a problem in recording: if the recording levels were set for paa and normal speakers, then the voices of the esophageal speakers, particularly in the production of the softest, voiced /a/, might fall near the limit of the tape recorder's range; yet if the recording levels were set for typical esophageal speakers, then the vocal intensity of paa and normal speakers would overload the cassette deck and produce a distorted recording. To remedy this problem a 12 dB attenuator was constructed. This attenuator was utilized if a subject's performance produced a V.U. reading in excess of "0" in those tasks from which frequency and intensity measures would be derived. To provide 12 dB attenuation, the investigator disconnected the microphone from the left channel of the cassette deck, connected the microphone to the attenuator, and connected the attenuator to the left channel of the cassette deck.

A second intensity attenuation problem was presented by the fact that subject head movement towards or away from the microphone would produce variability in recorded vocal intensity. Since this variability could mimic perturbation of the fundamental, a mechanism for maintaining constant mouth-to-microphone distance was needed. A headset with an adjustable microphone boom was constructed. The adjustable boom allowed movement of the microphone through a range of six inches.

Beyond the utilization of the same equipment for all recording, a second means of maintaining controlled recording conditions (in particular, the known reference tone) was instituted. The most recent study comparing esophageal, tracheoesophageal (t-e), and normal voice (Robbins, et al., 1981) found that t-e speakers exhibited the greatest median intensities of the three groups. In
that study, the t-e speakers' median vocal intensity clustered around 80 dB; however, mouth-to-microphone distance was six inches. Since mouth-to-microphone distance in the present investigation was four inches, intensity values three to four dB greater were expected; therefore, the reference tone was established as 84.5 dBSPL.

To establish this reference tone reliably for each subject, a multi-step procedure was implemented. To adjust the cassette deck input to "0" V.U. equivalent to 84.5 dBSPL, the headset microphone was aligned with a Bruel and Kjaer (B & K) Impulse Sound Level Meter, Type 2209. The microphone and the sound level meter were placed directly in front of and equidistant from an Electro-Voice, Model SP8D speaker which was connected to a Hewlett Packard Audio Oscillator, Model 200 AB via a Hewlett Packard 350D Attenuator Set. The oscillator was set to produce a 1000 Hz tone. The intensity of this tone was adjusted with the attenuator so that the output of the speaker registered 84.5 dBSPL on the sound level meter. The microphone had been connected to the left channel input of the cassette deck. Once the sound level meter registered 84.5 dBSPL, the left channel V.U. meter was adjusted to show "0" V.U.

While this procedure established 84.5 dBSPL as the referent for "0" V.U., it also proved cumbersome and time consuming. In order to reduce the time and equipment involved in setting the reference tone, the oscillator and attenuator were disconnected from the speaker and connected to the left channel line-in of the cassette deck. Again, the intensity of the 1000 Hz signal was adjusted with the attenuator set until "0" V.U. registered on the cassette deck. Since "0" V.U. represented 84.5 dBSPL, the tone recorded directly from the attenuator also was 84.5 dBSPL. The settings on the oscillator were X10 and 100 for frequency and 20 for gain. The setting for the attenuator set was 20 dB. By re-
establishing these settings on the oscillator and attenuator and adjusting the record level of the cassette deck to register "0" V.U., the reference signal of 1000 Hz at 84.5 dBSPL could be set rapidly for every recording session. This procedure also obviated the need for a speaker and sound level meter for every session. The lengthier procedure involving use of the speaker and the sound level meter was utilized as a check of recording system integrity following each transportation of the system or component to an outside site or following battery replacement for the microphone.

An additional method of insuring "0" V.U. equivalent to 84.5 dBSPL was effected. When the record level of the cassette deck has been set to the desired value, the face of the deck's record-level gain control and the gain control knob were scribed so that when the two marks aligned, "0" V.U. would equal 84.5 dBSPL. This unsophisticated method produced good accuracy. When compared to the signal from the attenuator, use of the marks consistently placed the needle of the V.U. meter on the "0", but not consistently in its center. The error produced was less that +0.5 dB. This method, of itself, was not used to set the desired amplitude levels; rather, it served as another check of system integrity. If, with the marks aligned, the oscillator-attenuator combination failed to provide a signal of approximately "0" V.U., then unusual conditions were occurring.

Once the 1000 Hz tone of 84.5 dBSPL had been established, five to 60 seconds of this tone were recorded in order to provide a referent for later analysis. For a large group of subjects, a five second signal was utilized because this provided an adequate reference in terms of frequency and intensity. This five second tone proved cumbersome, however, in adjusting the gain of the level recorder used in subsequent data compilation because the tone duration
was too short. A 60-second tone replaced the five second tone to provide a longer signal. This proved to be more than was necessary and subsequently was replaced by a 30-second signal.

ALTERNATE SITE RECORDINGS: Recording sessions at Wichita State University were conducted by Cynthia Spillers, M.A., CCC/SP, a doctoral candidate at that institution. Ms. Spillers wrote her Master's thesis in alaryngeal voice at the Ohio State University. Subjects were solicited from laryngectomees attending the 1982 convention of the International Association of Laryngectomees, held in Wichita. Six subjects participated: one woman using the BSVP, two women using esophageal voice, and three men using esophageal voice. The investigator provided Ms. Spillers with a custom-built headset, microphone and attenuator identical to the equipment utilized at Ohio State University. TDK SA-C 60 cassettes were used in a Yamaha Tc-800 GL stereo cassette deck with Dolby. Due to equipment differences the 1000 Hz reference tone was recorded at 89.5 dBSPL, instead of 84.5 dBSPL. To insure that recording procedures were identical, Ms. Spillers received the same forms as used at Ohio State University and a videotape of a mock recording session was prepared and sent to her as a training tool.

Recording sessions in Newark, New Jersey, were done at the United Hospital's Department of Otolaryngology. Recording equipment was identical to that used at Ohio State University except that the reference tone was provided by an Aurex Neovox electrolarynx. This electrolarynx replaced the oscillator and attenuator set in order to provide greater ease of transportation to New Jersey. The Neovox provided a complex tone with a fundamental frequency of 75 Hz. With a microphone-to-vibrating-surface distance of one inch the peak
intensity of the Neovox tone was 105.5 dBSPL. All recording in New Jersey was performed by this investigator.

RECORDING PROCEDURES: Prior to recording the voice sample, each subject received a packet of forms: Participant Information Sheet/Consent Form, Background Information Questionnaire, the first paragraph of the "Rainbow Passage", and eleven "hVd" sentences - "h" + vowel + "d" words embedded in the sentence "I will say _____" (Appendices B, C, D and E respectively). The subjects completed the former two forms and, to familiarize themselves with material, read the latter two forms. Once this was accomplished each subject was seated in an audiometric booth (International/Acoustic Corporation, Type 403 ATR) and the microphone positioned four inches (10 cm) from the subject's lips.

The microphone was positioned by having the subject bite on a tongue depressor and hold it parallel to the floor. A clear plastic six inch (15 cm) ruler was placed on the depressor and a distance of four inches was measured from the margin of the subject's lips directly below the nasal filtrum. The microphone boom was adjusted to place the center of the microphone head at the four inch mark. The microphone's orientation to the subject was such that the microphone's head was sagittal to the head of the subject. All subjects completed the set of tasks in the same order: latency between the command to phonate and the production of audible phonation, consistency of phonation, duration of phonation of the vowel /a/, repetition of the syllable /da/, oral reading of the first paragraph of the "Rainbow Passage", oral reading of the eleven "hVd" sentences, and production of softest, voiced and loudest, comfortable /a/. The procedures employed during the recording session are found in Appendix F.
Since analysis of the spectral properties of voice was planned for the last three tasks, the prevention of distortion due to overloading the recorder with too intense a signal was necessary. As described, a method of recording the subject's voices at or below "0" V.U. through use of a 12 dB attenuator was implemented. This attenuation was employed for 12 normal laryngeal speakers, 10 paa speakers and 10 esophageal speakers.

**Physical Measures Methodology and Equipment**

Each subject's vocal performance underwent three forms of analysis to derive the dependent variable values for subsequent statistical operations. During the consistency task, each subject was observed to insure that phonation was attempted only once for each trial. This required visual as well as auditory monitoring and therefore, could not be derived from the audio recordings. The individual conducting the recording session noted the number of successful attempts to phonate and divided this by total trials (10) to provide a percentage value for consistency. A B & K Level Recorder, Type 2304, was utilized to obtain the following measures: latency between the command to phonate and the onset of audible phonation, prolongation of /a/ on one air intake, the number of /da/ syllables produced on one air intake, the amount of time required to read the first paragraph of the Rainbow Passage, the amount of silent time present in the preceding reading, and the peak intensities of three soft and three comfortably loud productions of /a/. A B&K High Resolution Signal Analyzer, Type 2033, was used to measure the first three formants and fundamental frequencies of the quantal vowels /i,a,u/ produced in the sentences "I will say h__d".

**LEVEL RECORDER PROCEDURES:** The B & K Level Recorder, Type 2304 provides a permanent graphic record of the durational and intensity charac-
teristics of acoustic signals. In the time domain, one second may be represented as one, three, ten, thirty, or 100 millimeters. In the intensity domain, a range of zero to fifty decibels is displayed and is relative to the level of reference signal utilized.

The TEAC A-350 cassette deck with setting of CrO₂ Dolby in and maximum gain for left channel output was connected directly to the level recorder, except when the subject's performance required use of the 12 dB attenuator. To compensate for the attenuation, each subject's voice was amplified 12 dB by connecting the cassette deck to a Hewlett-Packard Amplifier, Model 450A, set for 20 dB amplification. The signal from this amplifier proved to be slightly in excess of 20 dB; therefore the amplifier was connected to a Hewlett-Packard 350 D Attenuator Set, which was connected to the level recorder. The reference 1000 Hz signal was displayed on the level recorder and the level recorder gain control adjusted to place the reference signal at the 30 dB line for all subjects except those from Wichita State University and from Newark, New Jersey. Since these subjects had reference tones of 89.5 and 105.5 dBSPL respectively, the reference signal was adjusted so that the zero line approximated 55 dBSPL. This initial adjustment was completed without amplification. To derive 12 dB amplification the signal was passed through the amplifying mechanism and adjusted with the attenuator set to 12 dB higher than the original reference line.

By this procedure, all measures of intensity recorded on the level recorder shared a common 0 dB baseline of 54.5 dBSPL and all measures of intensity could be expressed in dBSPL by adding 54.5 to the value from the level recorder. Of equal or greater import, this procedure also insured that all temporal measure shared a common baseline.
The graphic level recorder was set so that stylus movement was 500 mm per second and the lowest frequency sampled was 20 Hz. For recording the reference signal, the speed was set at 3 mm per second and the gain adjusted as needed. Six measures were obtained from the graphic level trace and the procedures for each follow:

1. Latency: The subject's cue to phonate was the investigator saying "Now" and the response was to say "ah". The investigator observed the movement of the recording paper through the level recorder while listening to the voice sample and marked the recording paper at the initiation of "now" and "ah", for all five trials of the task. Level recorder speed was set at 30 mm per second. The cue and response were considered as initiated when the recorder level tracing rose at an angle of 80° or more above the baseline (Berlin, 1963). The amount of time which transpired between the initiation of the cue and the initiation of the response defined latency. This time was calculated by measuring distance along the baseline between the onset of "now" and the onset of "ah". This distance was measure to the nearest 0.5 mm using a clear plastic 150 mm ruler. The distance was converted to time by dividng by 30. The time for all five trials was summed and averaged to yield a mean latency value for each subject.

2. Prolongation of /a/: Since Berlin's data (1963) indicated that proficient esophageal speakers generally produced mean durations of /a/ in excess of 1.8 seconds, the temporal resolution of the latency task was not required and the level recorder speed was set at 10 mm per second. Again, the cue and response were "now" and "ah" and both were marked on the tracing paper. The criterion of at least one 80° rise of the tracing also was retained as defining the beginning of phonation. The cessation of phonation was defined as
a return of the tracing to the baseline for 0.2 seconds (2 mm) or longer (Berlin, 1963). Duration was defined as the time between the onset and cessation of voicing and was calculated by measuring the distance between the rise in the trace and its return to baseline. The 150 mm ruler was used and the distance was measured to the nearest 0.5 mm. The distance was converted to time (seconds) by dividing by 10. The sum for all five trails was averaged to yield a mean vowel duration value for each subject.

3. Syllable repetition: Similar to the previous two tasks the cue was "now;" however, the subject's response was to repeat "dah" as often as possible on one air intake. The initiation and cessation of the subject's performance was marked and the number of syllables produced was derived by counting the peaks traced on the recording paper (Berlin, 1963) and by listening to and counting the number of syllables produced. The mean performance of each subject was obtained by averaging over five trials.

For most subjects a recording speed of 10 mm per sec. provided an accurate record; however, four subjects articulated /da/ so rapidly that the speed of the level recorder was increased to 30 mm per sec. in order to provide clearer definition between syllable peaks.

4. Rate of oral reading: Level recorder speed for this task was 10 mm per sec. The investigator marked the recording paper at the beginning of each of the six sentences of the Rainbow Passage and at the end of the sixth sentence. Total reading time was derived by measuring the linear distance to the nearest 0.5 mm between the beginning and end of the passage and dividing that distance by ten. This yielded the number of seconds required for the reading. Dividing total seconds by 60 and then dividing total words in the pas-
sage (98) by the new value yielded a measurement of oral reading as words per minute.

5. Silent time in oral reading: For the purposes of this study silence was defined as tracings less than 5 dB above the baseline and 0.05 seconds (0.5 mm) in duration. These values were derived from the same graphic record as was used for rate of oral reading. Again, distance was measured to the closest 0.5 mm. The amount of silence present in the entire passage and in the second and third sentences individually was computed by dividing the distance involved by 10.

6. Mean intensity of soft and loud /a/: In this task each subject produced the vowel /a/ three times as quietly as possible while still maintaining voice and three times as loudly as was comfortable. The investigator marked each production of /a/ as the paper proceeded through the level recorder at a speed of 10 mm per sec. The peak intensity was identified as the greatest excursion of stylus above the baseline. The perpendicular distance from the baseline to the peak was measured to the closest 0.5 mm and this distance converted to dBSPL by adding the distance in millimeters to 54.5. A mean value for the three soft and three loud productions was calculated as well as the difference between the two means and the difference between the least intense and most intense productions.

For two subjects the production of loud /a/ was not recorded due to operator error. In order to obviate the removal of these subjects from the study, the entire graphic trace for each subject was inspected and the three most intense peaks for each subject were substituted for the missing productions.

HIGH RESOLUTION SIGNAL ANALYZER PROCEDURES: The B & K High Resolution Signal Analyzer, Type 2033, provides realtime fast Fourier trans-
form analysis of transient acoustic events such as speech. The acoustic signal is sampled and stored in a 10K memory (1K = 1024) and displayed on a CRT screen as a 4000 line spectrum. The frequency range of this instrument is 0 to 20,000 Hz with 11 selectable baseband frequencies. The intensity range of the display is 80 dB relative to the reference signal. The B & K 2033 may be used to average the spectral characteristics of a signal automatically or a manual scan of the signal may be performed. Measures are to the closest tenth of a decibel and hundredth of a Hertz.

Input to the B & K 2033 was provided by connecting the TEAC A-350 stereo cassette deck directly to the analyzer. Settings for the A-350 were CrO₂, Dolby in and maximum gain for left channel output. No additional amplification was provided for those voice samples recorded with 12 dB attenuation since the attenuation could be compensated for by simply adding 12 dB to the B & K 2033's intensity values. Similarly, no attempt was made to establish the reference tone to a value of 84.5 or 89.5 dB, although this would have placed all intensity readings in dBSPL. Rather, the instrument was placed in the scan averaging mode, input function - time, input - direct, input - attenuation 110 dB, Hanning weighting utilized and full scale frequency - 2000 Hz. The reference signal was analyzed and averaged over 145 spectra. The average frequency and intensity of the reference signal were recorded. The displayed intensity always exceeded the known intensity of the reference tone in dBSPL; the difference between the displayed value and the known value for each subject was subtracted from any displayed intensity value of a subject's voice to furnish intensity values in dBSPL.

The instrumentation for the measurement of the fundamental and first three formant frequencies for the three quantal vowels was identical to that
used for analysis of the reference tone except that two different full scale frequency ranges were used (500 Hz and 5000 Hz) and the analysis averaged over 1153 spectra. The full scale frequency range of 500 Hz was used in the identification of the fundamental and first formants. The exception to this use was when the value of F<sub>1</sub> exceeded 500 Hz. In this case only f<sub>0</sub> could be extracted during the analysis of the 500 Hz range. The full scale frequency range of 5K Hz was used principally in the identification of the second and third formants and occasionally in the identification of the first formant. A setting of 1153 spectra was utilized because this provided the maximum resolution of frequency and intensity in the scan average mode.

To place a vowel sample in memory for analysis, the subject's production of the syllable "h__d" was presented with the B & K 2033 set for single record and internal trigger. For the 500 Hz scale the instrument's time window exceeded the duration of the syllable; therefore, at the end of the signal, record stop was selected manually to prevent extraneous acoustic stimuli from contaminating the analysis. Additionally the scan average also was controlled manually so that the scan began at the onset of the syllable and terminated at the end of the syllable.

Once the scan averaging was complete, the B & K 2033 was shifted from input function - time, which displayed the temporal and intensity characteristics of the signal, to input function - instrument, which displayed the frequency and intensity characteristics of the signal within the specified frequency range. Once these were displayed the fundamental and formants were identified as a narrow spectral peak for the fundamental or as broader spectral maxima for the formants. The data provided by Peterson and Barney (1952), Sisty and Weinberg (1972) and Robbins, et al. (1982) were utilized to
supply limiting boundaries for the interpretation of fundamental and formant locations on the spectrum.

The procedure for locating formants employed a different definition of formant frequency than previous works. Use of the sound spectrograph has lead to the identification of formants by their midpoint frequencies regardless of whether the actual point of energy maximum occurred at the midpoint (Lieberman, 1977). Since the B & K 2033 display allowed actual observation of the energy peak, the frequency of this peak, not of the midpoint, was used. Where a formant had more than one peak of equal intensity, the peak of lowest frequency was chosen to represent the formant frequency.

**Equal-Appearing Interval Scale**

**CONSTRUCTION:** The equal-appearing interval scale technique serves three purposes. First, such a scale allows for the measurement of psychological phenomena, such as vocal acceptability, in a reliable, replicable manner. Second, the measurements obtained may be treated as interval in nature, which permits use of powerful statistical techniques, i.e. analysis of variance. Third, use of the equal-appearing interval scale allows a large number of samples to be evaluated along the same metric while exposing the subjects to each sample only once (Thurstone and Chave, 1929; Edwards, 1957; Silverman, 1977). This third point is integral in selecting equal-appearing interval scaling for this investigation. With a total sample size of 45 speakers for the subjects to evaluate, the time involved in training subjects to use the scale and then to evaluate all 45 samples was approximately 45 minutes. An alternative method of scaling, such as the method of paired comparisons, would have required 990 comparisons, thus requiring nearly 16 hours.
In concept equal-appearing interval scale technique is straightforward. The basic assumptions are that attitude of a population on a given topic is linear in nature, that people will be honest in reporting what point on the linear continuum reflects their opinion, that certain points on the scale can be identified so reliably as to act as landmarks, and that subjects can be trained to use these landmarks in making judgements along the scale (Thurstone and Chave, 1929). Given these assumptions, an investigator must delineate the continuum of the attitude in question, identify the landmarks, and train subjects to use the landmarks. This process for the current investigation required two separate groups of listeners. One group of listeners was used to delineate the continuum and to identify the landmarks. An entirely separate second group was trained in use of the scale constructed from the responses of the first group.

To provide listeners with a representative sample of each speaker's voice, the speech samples were two sentences in length. The two sentences were the second and third in the "Rainbow Passage" (Fairbanks, 1960). Previous use of the equal-appearing interval technique in alaryngeal voice research had used only sentence two (Bennett and Weinberg, 1973). Horii (1975) had demonstrated that both sentences possessed very high correlations with the entire first paragraph of this passage on frequency related vocal characteristics; therefore, both sentences comprised the sample in order to increase the representative nature of the sample.

In constructing the scale, 50 voice samples were prepared either from recordings already on file in the Ohio State University Speech and Hearing Clinic, or by recording volunteers under the same conditions used to record
the experimental voice samples for this investigation. The voices present in
the initial 50 recordings represented buccal, pharyngeal, esophageal, and
tracheoesophageal alaryngeal voices, as well as voices of individuals with nor-
mal laryngeal mechanisms, vocal nodules, vocal polyps, hyponasality, and vocal
abuse. Other communication disorders such as stuttering were not present in
any of the voice samples.

The 50 samples were arbitrarily assigned numbers of 1-50 and then were
arranged in random order using a random numbers table (Silverman, 1977).
These samples were dubbed from a Marantz SD1015 stereo cassette deck or
from a Magnecord 1022 reel-to-reel stereo tape recorder to a TEAC A-350
stereo cassette deck using TDK SA-C60 Super Alvyn audio cassettes. The
1000 Hz reference signal was set at 0 V.U. on the TEAC deck with settings
of CrO2, line-in, and Dolby in.

A total of 50 students volunteered for this phase of the study. The criteria
for acceptance as a listener were normal hearing (by self-report), English as
the native language, and lack of familiarity with alaryngeal voice. Familiarity
was defined as having a close friend, associate, or relative who was a laryn-
gectomee. Forty-eight of the 50 students were enrolled in the following
courses: Speech and Hearing 235, Communication 305, 505, and 811. The
remaining two listeners were volunteers from the student body at Ohio State
University. These two were used to bring total listeners to 50. The age and
sex characteristics of this group is depicted in Table 3-1.
TABLE 3-1. Characteristics of listeners in construction of acceptability scale.

<table>
<thead>
<tr>
<th></th>
<th>#</th>
<th>Age Range</th>
<th>( \bar{X} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Males</td>
<td>10</td>
<td>19-37</td>
<td>27</td>
</tr>
<tr>
<td>Females</td>
<td>40</td>
<td>19-55</td>
<td>25</td>
</tr>
<tr>
<td>Total</td>
<td>50</td>
<td>19-55</td>
<td>25.5</td>
</tr>
</tbody>
</table>

These listeners assessed the voices for acceptability either in a classroom or in the Speech and Hearing Science laboratory. The equality of listening conditions was assured in the following manner.

The voices were played in a free field using an Electro-Voice Baronet high fidelity speaker, Harmon-Kardon Citation B solid state stereo power amplifier, and TEAC A-350 stereo cassette deck. The listeners were seated so that, within a tolerance of \( \pm 3 \) dB, the signal reaching them was 70 dB SPL (Bennett and Weinberg, 1973). This was achieved by playing a recorded white noise signal and adjusting the gain of the cassette deck until the desired level registered on a sound level meter. Levels were measured in each setting using a B & K Impulse Level Meter, Type 2209. The settings for these instruments were 70 dB range and slow needle movement. The settings for the TEAC cassette deck were \( \text{CrO}_2 \), MIC/DIN, and Dolby in. Using this equipment, seats in each setting which registered 70 dB (\( \pm 3 \) dB) levels were identified and listeners were required to place themselves in these seats.

The listeners were provided with a form for marking their assessment of each subject's voice (Appendix G). The listeners also received recorded in-
structions from the tape. These were identical to the written instructions, except that one sentence was added at the end: "If you have any questions, lose your place, or require more time, just signal the monitor and he will stop the tape." Verbal instructions were also given:

"You will be listening to a variety of voices, some normal, some deviant. Use the 1-9 scale to grade these voices for acceptability. You are not required to use the whole scale, for example, some voices may never be a 1 or a 9. If you wish to change any rating, put an X through your first mark and circle the new rating. You may change a rating whenever you wish. We will listen to the first six voices to acquaint you with the task."

From the accumulated assessments of the 50 listeners, median scale values and interquartile ranges (Q scores) were calculated for each of the samples. The median values were used to identify voices whose scale values fell at approximately equal intervals across the range of voices. The Q scores served as indices of the consistency with which listeners assessed a given voice. Narrow Q scores indicated close agreement among listeners for any single voice sample (Thurstone and Chave, 1929; Edwards, 1957); therefore, small Q scores were desirable in those samples contributing to the final scale.

For samples 7, 14, 18, 25, 31, and 46 the seventy-fifth percentile was located within the first interval of the nine point ordinal scale; therefore these voices were removed from further analysis because their scale values and Q scores could not be interpolated reliably. Since these six voices were of the lowest acceptability and the scale was to be applied only to proficient speakers' voices, the loss of sensitivity of the scale in its lowest region was not important to this study. From the remaining 44 samples, five were selected as most nearly producing a five point equal-appearing interval scale. The mean
Q score and standard deviation for the 44 samples were 1.2814 and 0.45. The mean Q score for the five samples selected for scale construction was 1.268.

The diagram below depicts the relative positions of the five scale values on a five point scale with the proportion of the entire scale accounted for by each interval. Ideally each interval should fill the same proportion of the scale. While this was not achieved, the mean percentile difference between successive intervals was 1.64% with the greatest difference between any two intervals being 2.87%. Clearly these values closely approximated equal intervals.

<table>
<thead>
<tr>
<th></th>
<th>25.72%</th>
<th>23.26%</th>
<th>26.13%</th>
<th>24.9%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>

The relative linear relationships of the five values used in the equal-appearing interval and the percentage of the scale within each interval.

APPLICATION: Once the five point scale had been constructed, the next step was to train the second group of listeners in use of the scale. These listeners needed to be able to identify reliably scale voice one as the least acceptable voice and scale voice five as the most acceptable voice, make appropriate judgements on the intervening three voices, and relate the levels of acceptability of the scale voices to the experimental voice samples.

The subjects for this portion of the study were 26 undergraduate students enrolled in a course entitled "Anatomy and Physiology of the Auditory and Vestibular Mechanism". Using the same criteria as were applied to the first group of listeners, these subjects were unfamiliar with alaryngeal voice. All of
these students were women. The mean age was 22.9 years in a range of 19 to 40 years.

The subjects were trained in use of the scale by listening to the scale twice, progressing from least to most acceptable, then to a random ordering of each voice twice, followed by a presentation of the entire scale from voice one to five. To ensure that the listeners remained consistent while applying the scale to evaluate the experimental voices, following every fifth experimental voice, scale voices one, three, or five were identified and presented to the listeners. Thus throughout the session the listeners were provided with the two ends and midpoint of the scale.

This segment of the study occurred in a classroom setting and the same procedures were utilized for maintaining equivalency of listening conditions as were described in the section on Scale Construction. Directions to the subjects and the recording form used are found in Appendices H and I.

As a check on the reliability of the listeners in using the scale, each scale voice was presented twice randomly as an experimental voice. The listeners, therefore, believed that they were evaluating 55 voice samples. Since each scale voice appeared twice to 26 listeners, there were 260 observations of the accuracy of the listeners in using the scale. Of these 260 observations, 23 were in error, i.e. a listener failed to identify the scale voice with the correct scale value. Of these 23 errors, one listener was responsible for five (22% of all errors); therefore, she was removed from the study, lowering the total number of listeners to 25.
Summary

Within Chapter III the procedures employed in this study have been described in four sections: speaking subjects, recording instrumentation, measurement of vocal acoustic characteristics, and the equal-appearing interval scale. In the following chapter the results of this investigation are presented and discussed. In Chapter V the study is summarized and implications for future research are outlined.
CHAPTER IV
Results and Discussion

Within this chapter the results of the statistical analysis of the data and interpretation of the results are presented. The organization of this chapter follows a topic line similar to Chapter II: first, the measurement of vocal acceptability is discussed; second, the analysis of frequency related measures is presented, followed by intensity related measures, and the analysis of temporal measures of voice. The main purpose in completing the frequency, intensity, and temporal analyses was to identify the most statistically prominent measures in these areas so that these measures could be related to the perceptual variable, vocal acceptability. Hence, the final section of this chapter presents the results of a multiple regression analysis of the physical measures of voice against vocal acceptability with an interpretation of this analysis.

Acceptability

The acceptability of the speaking subjects as judged by the naive listeners is summarized in Table 4-1. Overall, the data in this table indicate that in terms of vocal acceptability, normal voices are clearly superior to both forms of alaryngeal voice, that there is little if any difference between
the two types of alaryngeal voice, and that, as a group, laryngeal female voices were judged to be less acceptable than their male counterparts.

TABLE 4-1. Means and standard deviations (S.D.) of acceptability by voice type and sex of speaker.

<table>
<thead>
<tr>
<th></th>
<th>Laryngeal</th>
<th>PAA</th>
<th>Esophageal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>X 4.632</td>
<td>1.912</td>
<td>2.06</td>
</tr>
<tr>
<td>S.D.</td>
<td>.26</td>
<td>.25</td>
<td>.25</td>
</tr>
<tr>
<td>Female</td>
<td>X 3.512</td>
<td>2.192</td>
<td>2.136</td>
</tr>
<tr>
<td>S.D.</td>
<td>.27</td>
<td>.23</td>
<td>.35</td>
</tr>
</tbody>
</table>

This interpretation assumes greater credence in light of the two-factor, repeated measures ANOVA (Kennedy, 1978) performed on the naive listeners' judgements (Table 4-2). These listeners were exposed to three types of voice (factor one) produced by male and female speakers (factor two). This yielded a six-cell table with 25 observations per cell. The mean voice by sex effect was computed for each listener and entered into the appropriate cell.

TABLE 4-2. Repeated Measures ANOVA summary table for acceptability of voice by voice type and sex of speaker.

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>EMS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Voice (A)</td>
<td>2</td>
<td>127.9636</td>
<td>63.7818</td>
<td>E+SA+A</td>
<td>615.061*</td>
</tr>
<tr>
<td>Sex (B)</td>
<td>1</td>
<td>3.588</td>
<td>3.588</td>
<td>E+SB+B</td>
<td>49.445*</td>
</tr>
<tr>
<td>Subj. (S)</td>
<td>24</td>
<td>5.5631</td>
<td>0.2318</td>
<td>E+S</td>
<td>-</td>
</tr>
<tr>
<td>Voice X Sex (AB)</td>
<td>2</td>
<td>12.3321</td>
<td>6.166</td>
<td>E+SAB+AB</td>
<td>69.186*</td>
</tr>
<tr>
<td>Subj. X Voice (SA)</td>
<td>48</td>
<td>4.9797</td>
<td>0.1037</td>
<td>E+SA</td>
<td>-</td>
</tr>
<tr>
<td>Subj. X Sex (SB)</td>
<td>24</td>
<td>1.7417</td>
<td>0.0726</td>
<td>E+SB</td>
<td>-</td>
</tr>
<tr>
<td>Subj. X Voice X Sex (SAB)</td>
<td>48</td>
<td>4.2779</td>
<td>0.0891</td>
<td>E+SAB</td>
<td>-</td>
</tr>
<tr>
<td>TOTAL</td>
<td>149</td>
<td>160.0464</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* p \leq .001 for Geisser-Greenhouse F-Test (Kennedy, 1978)
Due to the repeated measurements design, only three F-tests were possible: Voice Type, Sex, and Voice by Sex interaction. Even with the application of the Geisser-Greenhouse procedure, limiting degrees of freedom to 1 and 24, all three F-tests were significant. Eta-squared coefficients were derived for each of these three tests. The most prominent effect as measured by the Eta-squared value was Voice Type (0.797) followed by the interaction (0.077) and Sex (0.022). Because the Eta-squared coefficient provides a measure of the proportion of total variance accounted for by an effect (Kennedy, 1978), the role of Voice Type in affecting vocal acceptability is undisputable and substantively unsurprising. The absence of a significant voice effect would lead to a conclusion that naive listeners do not perceive a difference between nonpathological, laryngeal voice and alaryngeal voice. Of greater interest was whether the two forms of laryngeal voice produced differential effects. Since only two pairwise comparisons were desired to test whether listeners perceived a significant difference in the acceptability of pulmonary assisted alaryngeal (paa) voice and esophageal voice, Dunn's procedure (Kennedy, 1978) was utilized with $\alpha = .05$ and Geisser-Greenhouse degrees of freedom 1 and 24. This yielded a critical value of 0.218. The means by Voice Type were 4.072 for normals, 2.134 for paa speakers, and 2.098 for esophageal speakers. Application of the 0.218 critical distance to normal speakers and nearest alaryngeal group, paa speakers, of course resulted in a statistically significant difference. The difference between the two alaryngeal groups, however, is 0.036 which fails to achieve significance.

The interaction effect of Voice by Sex appears attributable to the difference between laryngeal male and laryngeal females. To identify the exact source of this interaction, a second set of pairwise comparisons were made;
however, since four comparisons were desired, Tukey's test ($\alpha = .05$, df 1,24) was used. Tukey's "honest significant difference" was 0.238. This procedure requires that comparisons be made first between the extreme means in an ordered set. The comparisons of interest were normal females to normal males (difference= 1.12), normal females to the nearest alaryngeal group, paa females (difference= 1.32), paa females to paa males (difference= 0.116), and esophageal females to esophageal males (difference= 0.076). The first comparison in the ordered set was normal females to paa females and the normal females were found to be significantly more acceptable than the closest alaryngeal group. In comparison with normal males, however, the normal females were significantly less acceptable. This produced the interaction effect because no significant difference occurred between paa or esophageal males and females.

This difference between normal females and males also underlies the significant sex difference in this analysis. The cause of the difference can be related directly to two of the normal female speakers. Both of these subjects were judged by the panel of three speech pathologists as presenting voices normal for age and sex; however, in the ears of listeners, these speakers received mean acceptability scores of 1.68 and 1.64. The other three normal female speakers received scores of 4.64, 4.68 and 4.96. Why this skewed distribution of scores occurred for clinically normal voices is not apparent. Review of the voice samples by this investigator failed to reveal what factor was acting to degrade acceptability. A possible explanation is that due to the limited sample size, the normal female sample suffered from sampling error. What is of import to this study is that inclusion of these voices insures common ground in acceptability between normal and alaryngeal groups; therefore,
common factors may be at work.

**Frequency**

Two two-factor multivariate analyses of variance (MANOVA) were used to assess group differences measured in the frequency domain. Multivariate procedures were selected for two reasons: first, because of the large number of dependent variables employed in this study, a series of univariate tests (i.e. ANOVA) posed a threat to the control over the level of confidence and, therefore, to the control over Type I error. Second, multivariate methods allow for a broader perspective in data analysis and provide opportunities for identifying relationships between dependent variables, such that broader, theoretical constructs may become apparent.

The possibility of identifying more encompassing factors through application of multivariate techniques holds promise for the primary intent of this study, the identification of factors affecting vocal acceptability. The purpose in applying the four MANOVAs described in this chapter is to screen the physical measures of voice which have been obtained and to identify the most potent measures for inclusion in a multiple regression analysis against acceptability. The MANOVA technique, however, allows for not only the elimination of apparently weak factors, but also the consolidation of apparently potent factors into a single composite variable, Y (Tatsuoka, 1971). When a significant multivariate result is obtained, and the dependent variables involved lend themselves to the inference of higher order construct, an index of the higher order construct may be derived for each subject. The process involves converting each subject's performance on each dependent variable to a standardized score and multiplying each standardized score by the structure coef-
ficient for that variable, then summing the result. The SPSS MANOVA routine provides both structure coefficients and standardized scores (Hull and Nie, 1981).

The first MANOVA of frequency related data dealt with the ratio of the first formant to the second formant for each of the quantal vowels, /i,a,u/, with the fundamental frequency of the vowel /a/, and with the range in fundamental frequency among each subject's production of the quantal vowels. The anticipated results were that no significant differences in formant ratios would occur and that the alaryngeal groups would differ from the laryngeal group in terms of measures of fundamental frequency.

Tables 4-3 through 4-7 summarize the results of this MANOVA. A significant interaction between Voice Type and Sex occurred and the univariate F-tests for $f_o/a$, range of $f_o$, and for $F_2/F_1/u$ ratio were also significant. Despite the one significant formant ratio F-test, the locus of the difference in groups appears limited to measures of the vocal fundamental. Perusal of Table 4-7 reveals that in terms of both standardized weights and structure coefficients, the range of $f_o$ and $f_o/a$ account for the bulk of the difference.

Moving from the interaction to each of the main effects shows that significant multivariate results were obtained for both Sex and Voice Type (Pillai's V converted to F-tests of 3.6, d.f. 10,72, and 5.9, d.f. 5,35 respectively); but the univariate tests were significant only for $f_o/a$ and range of $f_o$, and the structure coefficients in both cases strongly emphasized the contribution of these two factors, to the exclusion of the formant ratios. The overall conclusion drawn from this analysis with respect to selecting factors possibly affecting vocal acceptability is that the two factors measuring aspects of fundamental frequency production are much more likely to be active in
TABLE 4-3. Means and standard deviations of $F_2:F_1$ for the Quantal vowels, for range of $f_o$ for the Quantal vowels, and for $f_o/a$ by voice and sex.

<table>
<thead>
<tr>
<th></th>
<th>$F_2:F_1$</th>
<th>$u$</th>
<th>$R_f_o$</th>
<th>$f_o/a$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$i$</td>
<td>$a$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normal</td>
<td>M 9.23 (1.64)</td>
<td>1.50 (0.21)</td>
<td>3.70 (1.01)</td>
<td>19.50 (7.03)</td>
</tr>
<tr>
<td></td>
<td>F 8.60 (2.01)</td>
<td>1.52 (0.29)</td>
<td>2.86 (1.04)</td>
<td>43.20 (18.46)</td>
</tr>
<tr>
<td>PAA</td>
<td>M 7.39 (1.15)</td>
<td>1.56 (0.18)</td>
<td>2.60 (0.65)</td>
<td>14.70 (9.19)</td>
</tr>
<tr>
<td></td>
<td>F 8.20 (2.47)</td>
<td>1.48 (0.11)</td>
<td>3.44 (0.40)</td>
<td>31.00 (9.14)</td>
</tr>
<tr>
<td>Eso</td>
<td>M 8.18 (1.13)</td>
<td>1.55 (0.20)</td>
<td>3.31 (0.56)</td>
<td>25.70 (15.26)</td>
</tr>
<tr>
<td></td>
<td>F 9.24 (1.86)</td>
<td>1.46 (0.18)</td>
<td>3.00 (1.13)</td>
<td>18.80 (10.45)</td>
</tr>
</tbody>
</table>
TABLE 4-4. MANOVA Summary Table for formant ratios and range of $f_o$ for the Quantal vowels and for $f_o /a/$.

<table>
<thead>
<tr>
<th>Test</th>
<th>Value</th>
<th>$F$</th>
<th>Hypothesis D.F.</th>
<th>Error D.F.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pillai's</td>
<td>0.738</td>
<td>4.2158*</td>
<td>10</td>
<td>72</td>
</tr>
<tr>
<td>Hotelling's</td>
<td>1.5382</td>
<td>5.2299*</td>
<td>10</td>
<td>68</td>
</tr>
<tr>
<td>Wilk's</td>
<td>0.3565</td>
<td>4.7236*</td>
<td>10</td>
<td>70</td>
</tr>
<tr>
<td>Roy's</td>
<td>0.5725</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* $p \leq .05$

TABLE 4-5. Summary of Univariate F-Tests for $F_{2:F_1}$ of /i,a,u/ and range of $f_o$ across /i,a,u/ and $f_o /a/$. D.F. (2,39).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Hypothesis MS</th>
<th>Error MS</th>
<th>$F$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$F_{2:F_1} /i/$</td>
<td>2.78</td>
<td>2.61</td>
<td>1.06</td>
</tr>
<tr>
<td>$F_{2:F_1} /a/$</td>
<td>0.01</td>
<td>0.04</td>
<td>0.35</td>
</tr>
<tr>
<td>$F_{2:F_1} /u/$</td>
<td>2.46</td>
<td>0.70</td>
<td>3.53*</td>
</tr>
<tr>
<td>$f_o$</td>
<td>849.64</td>
<td>139.34</td>
<td>6.10*</td>
</tr>
<tr>
<td>$f_o /a/$</td>
<td>8840.62</td>
<td>616.60</td>
<td>14.34*</td>
</tr>
</tbody>
</table>

* $p \leq .05$
TABLE 4-6. Weighted means for $F_2:F_1$ of /i,a,u/, range of $f_o$ across /i,a,u/ and $f_o/a$ by voice and sex.

<table>
<thead>
<tr>
<th></th>
<th>$F_2:F_1$</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>/i/</td>
<td>/a/</td>
<td>/u/</td>
<td>$Rf_o$</td>
<td>$f_o/a$</td>
</tr>
<tr>
<td>Normal</td>
<td>M</td>
<td>9.23</td>
<td>1.50</td>
<td>3.70</td>
<td>19.50</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>8.60</td>
<td>1.52</td>
<td>2.86</td>
<td>43.20</td>
</tr>
<tr>
<td>PAA</td>
<td>M</td>
<td>7.39</td>
<td>1.56</td>
<td>2.60</td>
<td>14.70</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>8.20</td>
<td>1.48</td>
<td>3.44</td>
<td>31.00</td>
</tr>
<tr>
<td>Eso</td>
<td>M</td>
<td>8.18</td>
<td>1.55</td>
<td>3.3</td>
<td>25.70</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>9.24</td>
<td>1.46</td>
<td>3.00</td>
<td>18.80</td>
</tr>
</tbody>
</table>

TABLE 4-7. Standardized weights and structure coefficients for $F_2:F_1$ of /i,a,u/, range of $f_o$ across /i,a,u/ and $f_o/a$ by voice and sex.

<table>
<thead>
<tr>
<th></th>
<th>$F_2:F_1$</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>/i/</td>
<td>/a/</td>
<td>/u/</td>
<td>$Rf_o$</td>
<td>$f_o/a$</td>
</tr>
<tr>
<td>Standardized Wts.</td>
<td>-0.296</td>
<td>0.394</td>
<td>-0.249</td>
<td>0.655</td>
<td>0.772</td>
</tr>
<tr>
<td>Structure Coeff.</td>
<td>-0.197</td>
<td>0.111</td>
<td>-0.166</td>
<td>0.436</td>
<td>0.740</td>
</tr>
</tbody>
</table>
TABLE 4-8. Means and standard deviations of the first two formants of the Quantal vowels by voice type and sex.

<table>
<thead>
<tr>
<th>Speakers</th>
<th>/i/F₁</th>
<th>/a/F₁</th>
<th>/u/F₁</th>
<th>/i/F₂</th>
<th>/a/F₂</th>
<th>/u/F₂</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>272.0(47.17)</td>
<td>778.8(104.1)</td>
<td>288.1(56.66)</td>
<td>2419.9(135)</td>
<td>1155.0(122)</td>
<td>1026.3(201)</td>
</tr>
<tr>
<td>F</td>
<td>326.0(59.02)</td>
<td>875.0(159.04)</td>
<td>371.2(49.59)</td>
<td>2717.4(216)</td>
<td>1282.6(196)</td>
<td>1037.4(291)</td>
</tr>
<tr>
<td>PAA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>323.5(56.23)</td>
<td>777.3(132.79)</td>
<td>392.3(70.04)</td>
<td>2342.5(246)</td>
<td>1198.6(185)</td>
<td>985.2(129)</td>
</tr>
<tr>
<td>F</td>
<td>355.8(74.20)</td>
<td>881.0(70.21)</td>
<td>393.0(57.37)</td>
<td>2797.6(447)</td>
<td>1290.4(90)</td>
<td>1365.0(270)</td>
</tr>
<tr>
<td>Eso</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>315.5(38.90)</td>
<td>795.8(187.78)</td>
<td>389.8(52.95)</td>
<td>2502.2(210)</td>
<td>1218.6(162)</td>
<td>1275.0(213)</td>
</tr>
<tr>
<td>F</td>
<td>329.4(64.78)</td>
<td>912.6(90.1)</td>
<td>380.8(33.74)</td>
<td>2955.0(160)</td>
<td>1332.4(169)</td>
<td>1159.0(445)</td>
</tr>
</tbody>
</table>
TABLE 4-9. MANOVA Summary Table for first two formants of Quantal vowels by sex.

<table>
<thead>
<tr>
<th>Test</th>
<th>Value</th>
<th>F</th>
<th>Hypothesis D.F.</th>
<th>Error D.F.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pillai's</td>
<td>0.5314</td>
<td>6.43*</td>
<td>6</td>
<td>34</td>
</tr>
<tr>
<td>Hotelling's</td>
<td>1.1340</td>
<td>6.43*</td>
<td>6</td>
<td>34</td>
</tr>
<tr>
<td>Wilk's</td>
<td>0.4686</td>
<td>6.43*</td>
<td>6</td>
<td>34</td>
</tr>
<tr>
<td>Roy's</td>
<td>0.5314</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* p ≤ .001

TABLE 4-10. Summary of Univariate F-Tests for first two formants of Quantal vowels by sex D.F. (1,39).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Hypothesis MS</th>
<th>Error MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>/I/ F1</td>
<td>11155.60</td>
<td>2944.60</td>
<td>3.79</td>
</tr>
<tr>
<td>/A/ F1</td>
<td>111433.21</td>
<td>18640.06</td>
<td>5.98*</td>
</tr>
<tr>
<td>/U/ F1</td>
<td>6216.71</td>
<td>3226.42</td>
<td>1.93</td>
</tr>
<tr>
<td>/I/ F2</td>
<td>1614432.40</td>
<td>56467.42</td>
<td>28.59*</td>
</tr>
<tr>
<td>/A/ F2</td>
<td>123358.04</td>
<td>25071.86</td>
<td>4.92*</td>
</tr>
<tr>
<td>/U/ F2</td>
<td>83966.68</td>
<td>60180.28</td>
<td>1.40</td>
</tr>
</tbody>
</table>

* p ≤ .05

TABLE 4-11. MANOVA Summary Table for first two formants of the Quantal vowels by voice type.

<table>
<thead>
<tr>
<th>Test</th>
<th>Value</th>
<th>F</th>
<th>Hypothesis D.F.</th>
<th>Error D.F.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pillai's</td>
<td>0.5306</td>
<td>2.1*</td>
<td>12</td>
<td>70</td>
</tr>
<tr>
<td>Hotelling's</td>
<td>0.7704</td>
<td>2.1*</td>
<td>12</td>
<td>66</td>
</tr>
<tr>
<td>Wilk's</td>
<td>0.5304</td>
<td>2.1*</td>
<td>12</td>
<td>68</td>
</tr>
<tr>
<td>Roy's</td>
<td>0.3622</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* p ≤ .05
determining acceptability, than are the formant ratios. This conclusion is in concert with the previous research into alaryngeal voice production; therefore, the range of \( f_0 \) for the quantal vowels and \( f_0/a/ \) are selected for inclusion in the multiple regression analysis.

The second MANOVA on frequency data was performed on the location of the peak frequency for the first two formants for each of the quantal vowels. While the pertinent literature indicated that no difference in formant ratios should have occurred, the same literature indicated that an upward shift in formant frequency should be expected. Two significant results were foreseen: 1. significant differences based on Voice Type founded in both alaryngeal groups demonstrating heightened formant frequencies vis a vis the laryngeal group; 2. significant differences based on Sex since women consistently have been shown to produce higher formant frequencies than men. The data assembled in Table 4-8 seem to bear out these expectations. The results of the MANOVA provide additional support in that there was no significant interaction effect, and there were significant main effects for voice and sex.

Referring back to Table 4-8, comparison of the means between sexes within voice type conforms to the prediction that women produce higher formants. Such a finding is not original, nor of great interest to this study.

Of greater interest is whether the expected elevation of formant frequencies occurred in the alaryngeal speakers. Again inspection of Table 4-8 bears out the expectation; but the question arises whether the differences are significant. The table of univariate results (Table 4-12) reveals only one area of difference across Voice Types, the production of \( F_1/u/ \). The structure coefficient associated with the first formant of /u/ also indicates that while it is the crux of the significant difference among Voice Types, it did not
stand alone. Y values on group means by voice were calculated: normal voice, Y= 0.788, paa voice, Y= -0.158, and esophageal voice, Y= -0.653. A clear distinction among groups appears based on these Y-scores and this distinction appears to be consistent with the available literature. The esophageal speakers show the greatest displacement from normal values because they have had their larynges removed and they alter their articulatory patterns to inject air into the esophagus. The paa speakers have also undergone laryngectomy; however, injection of air into the esophagus is unnecessary.

TABLE 4-12. Summary of Univariate F-Tests for first two formants of Quantal vowels by voice type D.F. (2,39).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Hypothesis MS</th>
<th>Error MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>/i/ F_1</td>
<td>7668.27</td>
<td>2944.59</td>
<td>2.60</td>
</tr>
<tr>
<td>/a/ F_1</td>
<td>2733.76</td>
<td>18640.06</td>
<td>0.14</td>
</tr>
<tr>
<td>/u/ F_1</td>
<td>27404.69</td>
<td>3226.42</td>
<td>8.49*</td>
</tr>
<tr>
<td>/i/ F_2</td>
<td>109630.07</td>
<td>56467.42</td>
<td>1.94</td>
</tr>
<tr>
<td>/a/ F_2</td>
<td>13077.22</td>
<td>25071.86</td>
<td>0.52</td>
</tr>
<tr>
<td>/u/ F_2</td>
<td>161933.09</td>
<td>60180.28</td>
<td>2.69</td>
</tr>
</tbody>
</table>

* p ≤ .001

Utilizing the structure coefficients and standardized dependent variables, Y scores for each subject based on formant production were calculated. These scores constituted the third set of variables for the multiple regression analysis.

Intensity

The analysis of intensity related dependent variables follows a course similar to that just encountered for frequency related measures; however, only one MANOVA was completed. Table 4-13 displays the means and standard deviations of the variables associated with vocal intensity. Since the
esophageal speakers lack the pulmonary air support to produce voice with strong intensity, it was anticipated that these speakers would produce, on the average, the least intense /a/ (XS), and the weakest most intense /a/ (XL). In terms of range of intensity from least to most intense production (LTM), the expected result was that laryngeal speakers would produce the greatest range followed by paa speakers. The variable range from mean soft /a/ to mean loud /a/ (XR) was used as a measure of a speaker's ability to control vocal intensity. It was assumed that the greater control a speaker had over voice, the closer XR would approach LTM, since the dispersion around XS and XL would be small. Since laryngeal speakers possess a more precisely controlled vocal mechanism, the expected outcome was that they would demonstrate the highest performance on this measure. Again paa speakers were expected to occupy a midpoint because they lacked a larynx, but retained control over the air supply for speech. A final assumption was that within each group, the male speakers would produce more intense voice than the female speakers by virtue of a larger air reservoir.

The data tabulated in 4-13 do not supply unequivocal support for the foregoing assumptions. Nor do the MANOVA results tabulated in 4-14; an interaction of voice by sex for intensity had not been predicted.

<table>
<thead>
<tr>
<th></th>
<th>XS</th>
<th>XL</th>
<th>LTM</th>
<th>XR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal M</td>
<td>80.08(6.35)</td>
<td>99.32(5.41)</td>
<td>22.40(8.60)</td>
<td>3.16(1.49)</td>
</tr>
<tr>
<td></td>
<td>F 76.17(8.48)</td>
<td>103.23(13.9)</td>
<td>34.00(11.29)</td>
<td>6.94(3.89)</td>
</tr>
<tr>
<td>PAA M</td>
<td>83.18(8.42)</td>
<td>93.84(6.55)</td>
<td>17.20(9.06)</td>
<td>5.17(3.14)</td>
</tr>
<tr>
<td></td>
<td>F 73.64(9.19)</td>
<td>85.33(6.83)</td>
<td>16.90(6.54)</td>
<td>4.222(1.44)</td>
</tr>
<tr>
<td>Eso M</td>
<td>75.88(6.35)</td>
<td>85.62(4.91)</td>
<td>14.85(5.12)</td>
<td>5.93(1.83)</td>
</tr>
<tr>
<td></td>
<td>F 76.29(9.15)</td>
<td>91.37(7.42)</td>
<td>18.90(12.12)</td>
<td>3.76(2.21)</td>
</tr>
</tbody>
</table>
TABLE 4-14. MANOVA Summary Table for effects of intensity factors by voice type and sex.

<table>
<thead>
<tr>
<th>Test</th>
<th>Value</th>
<th>F</th>
<th>Hypothesis D.F.</th>
<th>Error D.F.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pillai's</td>
<td>0.4402</td>
<td>2.61*</td>
<td>8</td>
<td>74</td>
</tr>
<tr>
<td>Hotelling's</td>
<td>0.5677</td>
<td>2.48*</td>
<td>8</td>
<td>70</td>
</tr>
<tr>
<td>Wilk's</td>
<td>0.6074</td>
<td>2.55*</td>
<td>8</td>
<td>72</td>
</tr>
<tr>
<td>Roy's</td>
<td>0.2477</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* p _ .05

TABLE 4-15. Standardized weights and structure coefficients for XS, XL, LTM, and XR by voice type and sex.

<table>
<thead>
<tr>
<th>XS</th>
<th>XL</th>
<th>LTM</th>
<th>XR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1stRt.*</td>
<td>2ndRt.</td>
<td>1stRt.</td>
</tr>
<tr>
<td>Stand.Wts.</td>
<td>-1.75</td>
<td>-1.29</td>
<td>1.24</td>
</tr>
<tr>
<td>Struct.Coeff.</td>
<td>-0.07</td>
<td>-0.54</td>
<td>-0.36</td>
</tr>
</tbody>
</table>

* Rt. - Root

TABLE 4-16. Weighted means (WX) of XS, XL, LTM, and XR by voice type and sex.

<table>
<thead>
<tr>
<th></th>
<th>WX of XS</th>
<th>WX of XL</th>
<th>WX of LTM</th>
<th>WX of XR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal M</td>
<td>80.08</td>
<td>99.32</td>
<td>22.40</td>
<td>3.16</td>
</tr>
<tr>
<td>F</td>
<td>76.17</td>
<td>103.23</td>
<td>34.00</td>
<td>6.94</td>
</tr>
<tr>
<td>PAA M</td>
<td>83.18</td>
<td>93.85</td>
<td>17.20</td>
<td>5.17</td>
</tr>
<tr>
<td>F</td>
<td>73.64</td>
<td>85.33</td>
<td>16.90</td>
<td>4.62</td>
</tr>
<tr>
<td>Eso M</td>
<td>75.88</td>
<td>85.62</td>
<td>14.85</td>
<td>5.93</td>
</tr>
<tr>
<td>F</td>
<td>76.29</td>
<td>91.37</td>
<td>18.90</td>
<td>3.76</td>
</tr>
</tbody>
</table>
TABLE 4-17. Summary of Univariate F-Tests for intensity factors by voice and sex  D.F. (2,39).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Hypothesis MS</th>
<th>Error MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>XS</td>
<td>82.98</td>
<td>59.55</td>
<td>1.39</td>
</tr>
<tr>
<td>XL</td>
<td>200.77</td>
<td>52.44</td>
<td>3.83*</td>
</tr>
<tr>
<td>LTM</td>
<td>120.85</td>
<td>74.55</td>
<td>1.62</td>
</tr>
<tr>
<td>XR</td>
<td>31.54</td>
<td>5.83</td>
<td>5.41*</td>
</tr>
</tbody>
</table>

* p ≤ .05

There were two significant roots (root one, F= 2.55 with d.f. 8 and 72; root two, F= 2.9, with d.f. 3 and 36.5). The structure coefficients for these two roots are found in Table 4-17. Since two roots were significant, two sets of Y scores for each subject were possible. Y scores on means of Voice by Sex were computed and plotted graphically to facilitate interpretation (Figure 4-1).

Figure 4-1. Y score means for roots 1 and 2 of MANOVA for intensity data plotted graphically.
Unfortunately an omnibus interpretation was not forthcoming to explain the dispersion of means through three quadrants; therefore an explanation based on the univariate F-tests was derived. Table 4-17 displays the F-tests for each of the four intensity dependent variables. Two tests were significant, $\overline{XL}$ and $\overline{XR}$. For both of these measures, the means in Table 4-13 reveal the source of the interaction. For $\overline{XL}$, contrary to expectations, laryngeal and esophageal females produced more intense voice than their male counterparts; but the reverse held true for paa speakers. To test whether the differences within each Voice Type were significant, Scheffe's method was used for follow-up testing. On degrees of freedom 1 and 28, the resultant F-tests were 0.97 for laryngeal speakers, 4.60 for paa speakers, and 2.10 for esophageal speakers. Of these only paa speakers exhibited a significant difference at the .05 level of confidence, a finding which identified the locus of the interaction effect.

To find the cause of the interaction for $\overline{XR}$, Scheffe's method was again applied to differences between sexes in each voice type. On 1 and 28 degrees of freedom, the comparison between normal speakers was significant ($F=8.17$) with $\alpha=.05$; however, neither the difference between paa speakers ($F=0.54$), nor the difference between esophageal speakers ($F=2.69$) were significant.

Since the purpose of this analysis was to screen dependent variables in order to derive a manageable number of variables for the multiple regression analysis, the main effects of the MANOVA were inspected to determine if both $\overline{XR}$ and $\overline{XL}$ should be retained or only one selected. There was no significant main effect for Sex; however, there was for Voice (Pillai's $V$ converted to $F=4.44$ on 8 and 74 degrees of freedom). Of the four univariate tests only two were significant: LTM, $F=6.24$ (d.f. 2, 39) and $\overline{XL}$, $F=13.15$ (d.f. 2, 39).
As LTM had not appeared as a prominent factor in the interaction, efforts were not directed at identifying its role in the voice type main effect. The role of XL was of interest; therefore, three comparisons of Voice Type group means were made using Tukey's procedure. The comparisons were laryngeal speakers (X = 101.27) to paa speakers (X = 89.59) and to esophageal speakers (X = 88.49), and a comparison of the two alaryngeal means. Tukey's critical distance was calculated to be 3.75; therefore both alaryngeal groups differed significantly from the laryngeal group, but not from each other.

Because XL appeared as a viable differentiator of speaker performance on two levels of this MANOVA, it was selected over XR for inclusion in the multiple regression analysis. This brought the total number of variables to four: f_o/a/, Rf_o, Y scores for formant data, and XL.

**Temporal Measures**

For temporal measures, as with formants, a higher order interpretation of results was possible based upon the voice by sex interaction which had only one significant root. Y scores were computed for each cell of the interaction. These are displayed in the diagram below.

![Diagram](Legend: see Figure 4-1.)

Y score means for voice by sex temporal measures

The overall interpretation of this effect is that the Y score measures a speaker's ability to rapidly initiate voice (LAT structure coefficient -0.305) and then to sustain voice (DUR and /da/ Rep. structure coefficients 0.403 and
.855) without repeated pauses (PT structure coefficient -.2). In other words the Y score based on these measures is an index of a speaker's overall proficiency in producing voice. Since the production of voice is the single factor separating all groups and is of intrinsic interest in this study, Y scores for each subject were calculated and used as the fifth variable in the multiple regression analysis.

### TABLE 4-18. Means and standard deviations of temporal measures by sex and voice.

<table>
<thead>
<tr>
<th></th>
<th>LAT.sec</th>
<th>DUR.sec</th>
<th>/da/Rep.X</th>
<th>Rate</th>
<th>Pause T. %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>M</td>
<td>.49(.20)</td>
<td>18.13(7.87)</td>
<td>84.7(36.66)</td>
<td>191.31(21.08)</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>.40(.06)</td>
<td>21.35(3.01)</td>
<td>154.82(35.68)</td>
<td>191.04(10.14)</td>
</tr>
<tr>
<td>PAA</td>
<td>M</td>
<td>.65(.21)</td>
<td>8.47(3.91)</td>
<td>31.63(10.90)</td>
<td>136.85(26.20)</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>.81(.21)</td>
<td>4.60(2.41)</td>
<td>17.4(10.1)</td>
<td>153.86(32.61)</td>
</tr>
<tr>
<td>Eso</td>
<td>M</td>
<td>.69(.21)</td>
<td>1.58(0.43)</td>
<td>12.14(8.4)</td>
<td>134.15(35.31)</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>.73(.24)</td>
<td>2.38(1.48)</td>
<td>29.2(34.25)</td>
<td>124.42(43.2)</td>
</tr>
</tbody>
</table>

### TABLE 4-19. MANOVA Summary Table for temporal measures by voice and sex.

<table>
<thead>
<tr>
<th>Test</th>
<th>Value</th>
<th>F</th>
<th>Hypothesis D.F.</th>
<th>Error D.F.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pillai's</td>
<td>0.4514</td>
<td>2.099*</td>
<td>10</td>
<td>72</td>
</tr>
<tr>
<td>Hotelling's</td>
<td>0.7256</td>
<td>2.467*</td>
<td>10</td>
<td>68</td>
</tr>
<tr>
<td>Wilk's</td>
<td>0.5681</td>
<td>2.287*</td>
<td>10</td>
<td>70</td>
</tr>
<tr>
<td>Roy's</td>
<td>0.4028</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* * p ≤ .05

<table>
<thead>
<tr>
<th>Variable</th>
<th>Hypothesis MS</th>
<th>Error MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>LAT</td>
<td>0.0505</td>
<td>0.0412</td>
<td>1.225</td>
</tr>
<tr>
<td>DUR</td>
<td>43.3199</td>
<td>19.6254</td>
<td>2.207</td>
</tr>
<tr>
<td>/da/ Rep.</td>
<td>6060.751</td>
<td>615.146</td>
<td>9.853*</td>
</tr>
<tr>
<td>Rate</td>
<td>612.843</td>
<td>859.7177</td>
<td>0.713</td>
</tr>
<tr>
<td>Pause T.</td>
<td>737.204</td>
<td>1407.9425</td>
<td>0.524</td>
</tr>
</tbody>
</table>

* p ≤ .05

### TABLE 4-21. Weighted means for temporal measures by voice and sex.

<table>
<thead>
<tr>
<th></th>
<th>LAT</th>
<th>DUR</th>
<th>/da/Rep.</th>
<th>Rate</th>
<th>Pause T.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>M</td>
<td>0.49</td>
<td>18.13</td>
<td>84.7</td>
<td>191.31</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>0.40</td>
<td>29.35</td>
<td>154.82</td>
<td>191.04</td>
</tr>
<tr>
<td>PAA</td>
<td>M</td>
<td>0.65</td>
<td>8.47</td>
<td>31.63</td>
<td>136.85</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>0.81</td>
<td>4.60</td>
<td>17.40</td>
<td>153.86</td>
</tr>
<tr>
<td>Eso</td>
<td>M</td>
<td>0.69</td>
<td>1.58</td>
<td>12.14</td>
<td>134.15</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>0.73</td>
<td>2.38</td>
<td>29.2</td>
<td>124.42</td>
</tr>
</tbody>
</table>

### TABLE 4-22. Standardized weights and structure coefficients for temporal measures by voice and sex.

<table>
<thead>
<tr>
<th></th>
<th>LAT</th>
<th>DUR</th>
<th>/da/Rep.</th>
<th>Rate</th>
<th>Pause T.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stand. Wts.</td>
<td>-0.243</td>
<td>0.122</td>
<td>0.871</td>
<td>-0.616</td>
<td>-0.198</td>
</tr>
<tr>
<td>Struct. Coef.</td>
<td>-0.305</td>
<td>0.403</td>
<td>-0.855</td>
<td>-0.150</td>
<td>-0.199</td>
</tr>
</tbody>
</table>
TABLE 4-23. MANOVA Summary Table for temporal measures by sex.

<table>
<thead>
<tr>
<th>Test</th>
<th>Value</th>
<th>F</th>
<th>Hypothesis D.F.</th>
<th>Error D.F.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pillai's</td>
<td>0.29</td>
<td>2.8568*</td>
<td>5</td>
<td>35</td>
</tr>
<tr>
<td>Hotelling's</td>
<td>0.408</td>
<td>2.8568*</td>
<td>5</td>
<td>35</td>
</tr>
<tr>
<td>Wilk's</td>
<td>0.710</td>
<td>2.8568*</td>
<td>5</td>
<td>35</td>
</tr>
<tr>
<td>Roy's</td>
<td>0.29</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* p ≤ .05


<table>
<thead>
<tr>
<th>Variable</th>
<th>Hypothesis MS</th>
<th>Error MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>LAT</td>
<td>0.0132</td>
<td>0.0412</td>
<td>0.320</td>
</tr>
<tr>
<td>DUR</td>
<td>0.025</td>
<td>19.625</td>
<td>0.001</td>
</tr>
<tr>
<td>/da/ Rep.</td>
<td>5913.004</td>
<td>615.146</td>
<td>9.612*</td>
</tr>
<tr>
<td>Rate</td>
<td>54.600</td>
<td>859.718</td>
<td>0.064</td>
</tr>
<tr>
<td>Pause T.</td>
<td>17528.178</td>
<td>14079.426</td>
<td>1.245</td>
</tr>
</tbody>
</table>

* p ≤ .005

TABLE 4-25. Table of weighted means for temporal measures by sex.

<table>
<thead>
<tr>
<th>LAT</th>
<th>DUR</th>
<th>/da/Rep.</th>
<th>Rate</th>
<th>Pause T.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>.61</td>
<td>9.394</td>
<td>42.82</td>
<td>154.103</td>
</tr>
<tr>
<td>Female</td>
<td>.646</td>
<td>9.444</td>
<td>67.14</td>
<td>156.44</td>
</tr>
</tbody>
</table>

TABLE 4-26. Standardized weights and structure coefficients for temporal measures by sex.

<table>
<thead>
<tr>
<th>LAT</th>
<th>DUR</th>
<th>/da/Rep.</th>
<th>Rate</th>
<th>Pause T.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stand. Wts.</td>
<td>0.274</td>
<td>-0.248</td>
<td>1.023</td>
<td>0.207</td>
</tr>
<tr>
<td>Struct.Coeff.</td>
<td>0.142</td>
<td>0.009</td>
<td>0.777</td>
<td>0.063</td>
</tr>
</tbody>
</table>
### TABLE 4-27. MANOVA Summary Table for temporal measures by voice type.

<table>
<thead>
<tr>
<th>Test</th>
<th>Value</th>
<th>F</th>
<th>Hypothesis D.F.</th>
<th>Error D.F.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pillai's</td>
<td>0.9846</td>
<td>6.98*</td>
<td>10</td>
<td>72</td>
</tr>
<tr>
<td>Hotelling's</td>
<td>5.0137</td>
<td>17.05*</td>
<td>10</td>
<td>68</td>
</tr>
<tr>
<td>Wilk's</td>
<td>0.1448</td>
<td>11.4*</td>
<td>10</td>
<td>70</td>
</tr>
<tr>
<td>Roy's</td>
<td>0.8284</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* p ≤ .05

### TABLE 4-28. Weighted means of temporal measures by voice type.

<table>
<thead>
<tr>
<th></th>
<th>Normal</th>
<th>PAA</th>
<th>Eso</th>
</tr>
</thead>
<tbody>
<tr>
<td>LAT</td>
<td>0.46</td>
<td>0.71</td>
<td>0.70</td>
</tr>
<tr>
<td>DUR</td>
<td>19.21</td>
<td>7.18</td>
<td>1.84</td>
</tr>
<tr>
<td>/da/Rep.</td>
<td>108.07</td>
<td>26.89</td>
<td>17.83</td>
</tr>
<tr>
<td>Rate</td>
<td>191.22</td>
<td>142.52</td>
<td>130.91</td>
</tr>
<tr>
<td>Pause T.</td>
<td>16.34</td>
<td>26.27</td>
<td>33.83</td>
</tr>
</tbody>
</table>

### TABLE 4-29. Standardized weights and structure coefficients for temporal measures by voice.

<table>
<thead>
<tr>
<th></th>
<th>LAT</th>
<th>DUR</th>
<th>/da/Rep.</th>
<th>Rate</th>
<th>Pause T.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stand. Wts.</td>
<td>-0.094</td>
<td>-0.603</td>
<td>-0.554</td>
<td>-0.26</td>
<td>-0.031</td>
</tr>
<tr>
<td>Struct. Coef.</td>
<td>0.266</td>
<td>-0.797</td>
<td>-0.795</td>
<td>-0.436</td>
<td>0.287</td>
</tr>
</tbody>
</table>
Multiple Regression Analysis

The purpose of this analysis is to attempt to determine those factors active in affecting vocal acceptability. From the preceding MANOVAs, three dependent variables, $\bar{X}_L$, $f_o/a/$, and $Rf_o$, and two composite variables, formant elevation and voice production proficiency, were selected as potent discriminators among the three groups of speakers. In addition to these five, two other variables were added to the multiple regression in order to control for effects on acceptability. Since the speakers for this study came not only from Central Ohio, but also from New Jersey and Kansas, the geographical location at which the speaker was recorded was entered as a three level factor (1 = Central Ohio, 2 = Kansas, 3 = New Jersey). Additionally, the ANOVA on the acceptability measure revealed a difference based on sex of the speaker; therefore this variable was entered into the multiple regression as a two level factor (1 = male, 2 = female).

The sequential ordering of variables in a multiple regression can affect the relative weights derived for each variable, such that a later appearing variable which has a strong correlation with an earlier appearing variable may appear as less important than it actually is. To control for this possibility, the SAS Stepwise Regression program with MAXR option was utilized. This program begins by selecting the single dependent variable with the highest correlation with the variable of interest, then the two variable combination with the highest correlation, and so on until all variables occur in a saturated model or until no further significant increase in the correlation can be achieved (Helwig and Council, 1979). In this manner, the order of variables is determined solely by their contribution to reducing unaccounted for variance.
Table 4-30 displays the order of the models developed by the MAXR option from the one variable model to the seven variable model. From a theoretical perspective the occurrence of the mean loudest, comfortable production of /a/ (XL) is gratifying. The intensity of voice is largely a function of efficient use of the pulmonary air stream; therefore, those speakers able to use most efficiently the pulmonary air stream to produce voice (i.e. normal, laryngeal speakers) should have produced the most acceptable voices. The inclusion of speaker's geographic location in the two variable model is not quite so gratifying. The most reasonable explanation of this model is that once listeners ascertain that the voice is normal, they next attend, not to other physical characteristics of the vocal signal, but to the differences in pronunciation and intonation which characterize regional dialects.

The three variable model re-introduces theoretical gratification. A large portion of Chapter II related the changes in vocal fundamental frequency subsequent to total laryngectomy. The inclusion of $f_o/a/ \text{ with a positive weight}$ indicates that as frequency rises so also does vocal acceptability. Since the normal voices in this study displayed the highest fundamental frequencies for each sex, and were judged as exhibiting the most acceptable voices, the appearance of $f_o/a/$ appears reasonable and is in keeping with the Shipp article (1967) which ascribed a significant role to vocal fundamental in distinguishing skilled use of esophageal voice.

The appearance of sex in the four variable model can be explained on two grounds: first, the substantial difference between normal laryngeal males and females with respect to acceptability suggested strongly that sex of the speaker played a role in determining vocal acceptability. The second reason is tied to the appearance of fundamental frequency in the preceding model.
TABLE 4-30. Multiple regression models of the correlation of the three dependent variables (XL, $f_{oa}$, $R_{fo}$), of the two composite variables (phonatory proficiency and formant elevation), and location and sex of speaker with vocal acceptability. (F-Tests compare adjacent models.)

<table>
<thead>
<tr>
<th>Model</th>
<th>Equation</th>
<th>$R^2$</th>
<th>$F$</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 factor</td>
<td>$AC = -4.5633 + 0.00078 \times XL$</td>
<td>0.3260</td>
<td>20.80</td>
<td>0.0001</td>
</tr>
<tr>
<td>2 factor</td>
<td>$AC = -3.8885 + -0.6025 \times Loc + 0.0008 \times XL$</td>
<td>0.4014</td>
<td>14.08</td>
<td>0.0001</td>
</tr>
<tr>
<td>3 factor</td>
<td>$AC = -3.7029 + -0.6025 \times Loc + 0.0007 \times XL + 0.0052 \times f_{oa}$</td>
<td>0.4279</td>
<td>10.22</td>
<td>0.0001</td>
</tr>
<tr>
<td>4 factor</td>
<td>$AC = -2.7750 + -0.5747 \times Loc + -0.6377 \times Sex + 0.0007 \times XL + 0.0085 \times f_{oa}$</td>
<td>0.4743</td>
<td>9.02</td>
<td>0.0001</td>
</tr>
<tr>
<td>5 factor</td>
<td>$AC = -2.7460 + -0.5171 \times Loc + -0.6467 \times Sex + 0.0006 \times XL + 0.0069 \times f_{oa} + 0.0002 \times Phon$</td>
<td>0.4801</td>
<td>7.20</td>
<td>0.0001</td>
</tr>
<tr>
<td>6 factor</td>
<td>$AC = -2.7320 + -0.4990 \times Loc + -0.6047 \times Sex + 0.0007 \times XL + 0.0064 \times f_{oa} + -0.00009 \times For + 0.0002 \times Phon$</td>
<td>0.4830</td>
<td>5.92</td>
<td>0.0002</td>
</tr>
<tr>
<td>7 factor</td>
<td>$AC = -2.6576 + -0.4942 \times Loc + -0.5916 \times Sex + 0.0006 \times XL + 0.0066 \times f_{oa} + -0.002 \times R_{fo} + -0.00009 \times For + 0.0002 \times Phon$</td>
<td>0.4833</td>
<td>4.94</td>
<td>0.0005</td>
</tr>
</tbody>
</table>
Increasing vocal fundamental frequency can improve vocal acceptability only to a limited degree. A mature male speaker producing a fundamental frequency of 250 Hz (double the expected value) would probably suffer a decrement in acceptability due to the perception of inappropriately high pitch. The inclusion of sex in the four variable model helps to delineate the degree of increase in $f_0$ which would have a positive effect on acceptability. Given that sex received a prominent, negative beta weight, the strongest reason for its inclusion in the model probably resides in the difference between laryngeal men and women in acceptability.

The first composite score, phonatory proficiency, appears in the five variable model. While the increase in R-squared was less than .01, the index of phonatory proficiency did produce a significant increase in the correlation. The sign of its beta weight is positive which supports the interpretation of this composite as a measure of the ability to initiate and maintain phonation, although the value of the weight itself is small. The explanation for this diminutive weight is that while the composite does measure phonatory proficiency, the groups of speakers were selected on a criterion of proficient voice production. The abilities represented in the composite score are too elementary in voice production to account for differences found among groups skilled in producing the respective voice types; therefore, a reduced beta weight resulted and the relatively late appearance of this composite score as a factor affecting acceptability occurred.

The second composite score, formant elevation, first produces a significant effect in the second-to-last, or six-factor model. Inclusion of this factor in an early model was not expected since the laryngectomized speaker compensates for the formant elevation by maintaining the integrity of the $F_2:F_1$ ratio,
thereby maintaining vowel intelligibility. As the addition of the composite representing formant elevation produced a minute, but significant, increase in R-square, it is evident that although the listener's perception of vowels is preserved by the laryngectomee's compensatory articulation, the formant elevation is perceived and detracts from the normalcy of the alaryngeal voice.

The final variable in the multiple regression analysis is range of the fundamental frequency ($R_f$) across the quantal vowels. Again a significant, if miniscule increment to R-square resulted. As a factor in influencing listeners' judgements of vocal acceptability, $R_f$ proves to be largely inconsequential. The negative sign of the beta weight runs contrary to the current literature which suggest depressed variability of $f_0$ within the alaryngeal population. The negative weight indicates that as variability of the fundamental increases, acceptability decreases, which seems to imply that alaryngeal speakers should demonstrate heightened acceptability scores because the alaryngeal phonatory mechanism is less adjustable than the laryngeal phonatory mechanism. A more plausible explanation is that listeners prefer consistency in speakers' voices and than undue variability is perceived as unpleasant.

In an attempt to gauge the validity of the composite scores, a second multiple regression analysis was completed. The composite scores were removed and the most prominent single variables from each of the MANOVAs plus location and sex were used. Table 4-31 summarizes this analysis. Two variables which were present in the two composite scores, /da/ Rep. and F₁/u/, show early prominence, suggesting that blending them with the other components of the composite for phonatory proficiency and formant elevation
<table>
<thead>
<tr>
<th>Model</th>
<th>Equation</th>
<th>R-square</th>
<th>F</th>
<th>p_</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 factor</td>
<td>AC = 2.1261 + .013(\text{da/Rep.})</td>
<td>0.2804</td>
<td>16.76</td>
<td>.0002</td>
</tr>
<tr>
<td>2 factor</td>
<td>AC = 4.4176 + 0.1082(\text{da/Rep.}) + -0.006(F_{1/u})</td>
<td>0.3779</td>
<td>12.76</td>
<td>.0001</td>
</tr>
<tr>
<td>3 factor</td>
<td>AC = 4.6396 + -0.4057(\text{Sex}) + 0.0119(\text{da/Rep.}) + -0.0053(F_{1/u})</td>
<td>0.339</td>
<td>9.07</td>
<td>.0001</td>
</tr>
<tr>
<td>4 factor</td>
<td>AC = 4.4528 + -0.511(\text{Sex}) + 0.0097(\text{da/Rep.}) + 0.0014(\text{XL}) + -0.0049(F_{1/u})</td>
<td>0.418</td>
<td>7.18</td>
<td>.0002</td>
</tr>
<tr>
<td>5 factor</td>
<td>AC = 4.3256 + -0.6169(\text{Sex}) + 0.0096(\text{da/Rep.}) + 0.0013(\text{XL}) + -0.0052(F_{1/u}) + 0.0005(f_{o/a})</td>
<td>0.431</td>
<td>5.91</td>
<td>.0004</td>
</tr>
<tr>
<td>6 factor</td>
<td>AC = 4.4631 + -0.6173(\text{Sex}) + -0.1674(\text{LOC}) + 0.0096(\text{da/Rep.}) + 0.0013(\text{XL}) + -0.0049(F_{1/u}) + 0.0047(f_{o/a})</td>
<td>0.4354</td>
<td>4.88</td>
<td>0.0009</td>
</tr>
</tbody>
</table>
smothered the effects of /da/ Rep. and F₁/u/ on acceptability. A more meaningful rationale is that these two variables were untempered by appearing out of context with other related measures and, therefore, were ascribed with a greater effect on acceptability than otherwise would have occurred. This rationale is born out by the increase in R-square which resulted from use of the composite scores and from the relatively late appearance of these scores in the analysis.

The final chapter of this investigation interprets these results from a clinical perspective and outlines implications for future research.
CHAPTER V
Summary and Conclusions

While the average clinician may have difficulty extracting clinically relevant information from the foregoing detailed statistical analysis, the clinical implications of these findings may be viewed as very powerful to those involved in larynectomyee rehabilitation. The efforts of this investigation have been directed at determining the nature of pulmonary assisted alaryngeal (paa) voice in light of normal voice and esophageal voice. While on specific measures paa voice may vary from proficiently produced esophageal voice, taken in overview, these two forms of alaryngeal voice are highly similar to each other in both the physical and perceptual domains. Such a conclusion necessarily implies that paa voice deviates substantially from normal laryngeal voice and the accumulated data confirm this implication.

If the two forms of alaryngeal voice are so alike, the clinician is led to question the advantage to the larynectomyee in using paa voice. Two factors not measured in this study bear upon this question. First, in the experience of this investigator and others (Rusnov, 1981), paa voice restoration procedures are successful in providing functional speech for approximately 70% of the clients treated in this manner. The recognized success rate for esophageal voice therapy is 10-20% lower; therefore paa voice therapy allows the return of vocal communication to a greater number of larynectomyees. Second, in the majority of cases, patients undergoing paa voice restoration leave the hospital speaking within four to six days after the shunt surgery. One paa
A speaker in this study had utilized her voice prosthesis for only three weeks prior to recording, yet still was judged as a proficient speaker. No esophageal speaker had utilized esophageal voice for less than six months prior to recording. In short it takes much longer to train proficient esophageal speech than to train proficient paa speech; therefore functional speech is restored more rapidly with paa voice.

An important proviso must be noted: all voice recordings in this study were obtained in a quiet environment, which is not the usual communicative setting for most speakers or listeners. The effect of a noisy environment upon the user of and listener to paa voice is unknown. Can the paa speaker adjust vocal output to compensate for environmental noise better than the esophageal speaker? The data in Robbins, et al. (1981) indicate that such an increase in vocal intensity is possible, while the results from this study indicate that such an increase may not be comfortable for the speaker. Additionally the effects of the paa speaker's compensatory efforts upon the listener is an open question. In terms of vocal physiology the paa speaker should be able to produce and sustain more intense speech than the esophageal speaker; hence, the paa speaker's intelligibility in noise should be better preserved. The experimental conditions enforced for this study supply no insights into this area, and thus constitute an important limitation on the generalizability of the study results.

With the exception of the Clark and Stemple article (1982) no available research into the nature of paa voice has addressed the effects of adverse speaking and listening conditions. Clark and Stemple's procedures recorded the speakers in quiet; then dubbed noise onto the audio tapes, obviating any measurement of speaker compensation for the added noise. Future research
into the effectiveness of paa voice should address this issue of speaking and listening under adverse conditions. Clark and Stemple make a valid point in proffering the use of recognized intelligibility tests for measuring the effects of noise; however, their procedure should be modified so that the speakers are recorded in various environments of known noise levels.

A second limitation of the present study deals with the listeners, young adult females, used to gauge acceptability. These judges were critical of post-menopausal female voices and of alaryngeal voice in general. Assuming that laryngectomees work and soicalize with individuals in their peer group (mid to late 50's or above), the question of how the older listener perceives alaryngeal voice is valid but unanswered.

**Conclusion**

In Chapter I three questions concerning paa voice were posed. The first question dealt with whether paa voice was perceptually more pleasing than esophageal voice. The answer to this question, as measured on an equal-appearing interval scale of vocal acceptability is that no significant differences were found. The second question inquired into the physical characteristics of voice underlying any difference in paa and esophageal acceptability. While some differences in single physical measures were identified, these had no evident impact on acceptability. The third question focused on a series of tasks traditionally used for documenting improvement in the acquisition of alaryngeal voice. The condensing of these tasks into a single composite score reflecting phonatory proficiency in the alaryngeal speaker, and the appearance of this composite relatively late in the multiple regression analysis indicates that these tasks are foundational to acquiring voice and are valid indices of progress in therapy.
BIBLIOGRAPHY


. Some statistical characteristics of voice fundamental frequency. Journal of Speech and Hearing Research, 1975, 18, 192-201.

. and Weinberg, B. Intelligibility characteristics of superior esophageal speech presented under various masking conditions. Journal of Speech and Hearing Research, 1975, 18, 413-419.


Shapiro, M. Personal communication. September 15, 1982.


APPENDIX A

SPEECH PATHOLOGIST'S EVALUATION
OF SPEAKER VOCAL PROFICIENCY
Appendix A

Please evaluate the following 45 voices along the following guidelines:

1. Laryngeal speakers: Would you enroll this speaker in voice therapy? (Circle Y or N). Does this voice sound normal to you? (Circle Y or N). All normal speakers are in their mid 50's or older.

2. Esophageal and puncture (use of valved prosthesis) speakers: Is this person proficient in using the form of alaryngeal speech presented? (Circle Y or N). Rate the speaker on this scale B.A.- Below Average, (Speech barely functional; not proficient) A. - Average (Proficient speaker, representative of typical esophageal/puncture speaker), A.A.- Above Average, (Very proficient, among the better esophageal/puncture voices you have heard), S. - Superior (One of best esophageal/puncture voices you have heard).

Voice 1 Esophageal, Male. Proficient? Y N Rate: B.A. A. A.A. S.
Voice 2 Esophageal, Female. Proficient? Y N Rate: B.A. A. A.A. S.
Voice 3 Laryngeal, Female. Therapy? Y N Normal? Y N
Voice 4 Skip
Voice 5 Laryngeal, Male. Therapy? Y N Normal? Y N
Scale Voice 1 Skip all scale voices (There are a total of 10)
Voice 6 Esophageal, Male. Proficient? Y N Rate: B.A. A. A.A. S.
Voice 7 Puncture, Male. Proficient? Y N Rate: B.A. A. A.A. S.
Voice 8 Skip
Voice 9 Esophageal, Female. Proficient? Y N Rate: B.A. A. A.A. S.
Voice 10 Laryngeal, Male. Proficient? Y N Normal? Y N
Scale Voice 3 Skip
Voice 11 Puncture, Male. Proficient? Y N Rate: B.A. A. A.A. S.
Voice 12 Puncture, Male. Proficient? Y N Rate: B.A. A. A.A. S.
Voice 13 Puncture, Male. Proficient? Y N Rate: B.A. A. A.A. S.
Voice 14 Skip
Voice 15 Puncture, Male. Proficient? Y N Rate: B.A. A. A.A. S.
Scale Voice 5  Skip
Voice 16  Puncture, Female.  Proficient? Y N Rate: B.A. A. A.A. S.
Voice 17  Skip
Voice 18  Puncture, Male.  Proficient? Y N Rate: B.A. A. A.A. S.
Voice 19  Laryngeal, Male.  Therapy? Y N Normal Y N
Voice 20  Laryngeal, Male.  Therapy? Y N Normal Y N
Scale Voice 1  Skip
Voice 21  Puncture, Female.  Proficient? Y N Rate: B.A. A. A.A. S.
Voice 22  Esophageal, Male.  Proficient? Y N Rate: B.A. A. A.A. S.
Voice 23  Laryngeal, Male.  Therapy? Y N Normal? Y N
Voice 24  Laryngeal, Male.  Therapy? Y N Normal? Y N
Voice 25  Laryngeal, Male.  Therapy? Y N Normal? Y N
Scale Voice 3  Skip
Voice 26  Skip
Voice 27  Esophageal, Female.  Proficient? Y N Rate: B.A. A. A.A. S.
Voice 28  Laryngeal, Female.  Therapy? Y N Normal? Y N
Voice 29  Laryngeal, Male.  Therapy? Y N Normal? Y N
Voice 30  Laryngeal, Male.  Therapy? Y N Normal? Y N
Scale Voice 5  Skip
Voice 31  Skip
Voice 32  Skip
Voice 33  Esophageal, Female.  Proficient? Y N Rate: B.A. A. A.A. S.
Voice 34  Skip
Voice 35  Puncture, Male.  Proficient? Y N Rate: B.A. A. A.A. S.
Scale Voice 1  Skip
Voice 36  Laryngeal, Female.  Therapy? Y N Normal? Y N
Voice 37  Esophageal, Male.  Proficient? Y N Rate: B.A. A. A.A. S.
Voice 38  Esophageal, Male.  Proficient? Y N Rate: B.A. A. A.A. S.
Voice 39  Skip
Voice 40  Puncture, Male.  Proficient? Y N  Rate:  B.A.  A. A.A. S.
Scale Voice 3  Skip
Voice 41  Esophageal, Male.  Proficient? Y N  Rate:  B.A.  A. A.A. S.
Voice 42  Skip
Voice 43  Puncture, Male.  Proficient? Y N  Rate:  B.A.  A. A.A. S.
Voice 44  Laryngeal, Female.  Therapy? Y N  Normal? Y N
Voice 45  Puncture, Male.  Proficient? Y N  Rate:  B.A.  A. A.A. S.
Scale Voice 5  Skip
Voice 46  Esophageal, Male.  Proficient? Y N  Rate:  B.A.  A. A.A. S.
Voice 47  Esophageal, Male.  Proficient? Y N  Rate:  B.A.  A. A.A. S.
Voice 48  Puncture, Female.  Proficient? Y N  Rate:  B.A.  A. A.A. S.
Voice 49  Esophageal, Male.  Proficient? Y N  Rate:  B.A.  A. A.A. S.
Voice 50  Puncture, Female.  Proficient? Y N  Rate:  B.A.  A. A.A. S.
Scale Voice 1  Skip
Voice 51  Laryngeal, Male.  Therapy? Y N  Normal? Y N
Voice 52  Esophageal, Male.  Proficient? Y N  Rate:  B.A.  A. A.A. S.
Voice 53  Laryngeal, Female.  Therapy? Y N  Normal? Y N
Voice 54  Esophageal, Female.  Proficient? Y N  Rate:  B.A.  A. A.A. S.
Voice 55  Puncture, Male.  Proficient? Y N  Rate:  B.A.  A. A.A. S.
Voice 56  Puncture, Female.  Proficient? Y N  Rate:  B.A.  A. A.A. S.
APPENDIX B

PARTICIPANT INFORMATION SHEET/
CONSENT FORM
First of all, thank you for volunteering for this study. The purpose of this form is simply to inform you about my study, what I hope to accomplish, and what you will be asked to do.

This study focuses on two recent advances in restoring voice following total laryngectomy, the Blom-Singer Voice Prosthesis and the Panje Voice Button. I will compare the voice produced using these two new devices with each other, with esophageal voice, and with normal laryngeal voice. Also, I will ask anonymous listeners to compare the various voices. I hope to identify possible advantages and disadvantages of the two new devices.

To do this, there is a variety of speech exercises which I will ask you to perform. In the first three tasks I simply want you to say "ahh" when I ask you to do so. I will be measuring how little delay you have in saying "ahh", how consistent you are, and for how long you can prolong "ahh" on one breath. Next I will ask you to say "dah" as often as you can on one breath. Following this, you will read a 98 word reading passage at a rate that is comfortable for you. The next exercise also requires reading, but this will be 11 four-word sentences. You will be asked to do all these exercises at a volume (loudness) that is comfortable and natural for you.

In the last task, I will ask you to say "ahh" 3 times as softly as possible while still voicing and not whispering and as loudly as is comfortable for you. If you have any questions or concerns, please ask me about them. I ran myself through the entire set of tasks in about 10 minutes, so there is plenty of time to answer your questions.

Consent

I consent to participating in a study entitled "A Comparison Study of Voice Proficiency and Vocal Characteristics of Tracheoesophageal, Esophageal, and Laryngeal Speakers". Dr. Sheila M. Goff or Michael D. Trudeau have explained the purpose of the study and procedure to be followed. Possible benefits have been described.

I acknowledge that I have had the opportunity to obtain additional information regarding the study and that any questions I have raised have been answered to my full satisfaction. Further, I understand that I am free to withdraw consent at any time and discontinue participation in the study without prejudice to myself. The information obtained from me will remain confidential and anonymous unless I specifically agree otherwise.

Finally, I acknowledge that I have read and fully understand the consent form. I have signed it freely and voluntarily and understand a copy is available upon request.

Date:________________________  Signed:__________________________________________

(Participant)

(Investigator/  (Site)  (Type)  (No.)
Authorized Rep.)
Background information

Do not sign your name on this form. We want to preserve your right to privacy.

Participant: ________________________ Site: ________________________ Age: ___________ yrs of education: ___________

Sex: ________________________ Date of original laryngectomy: ________________________

Date of prosthesis fitting: ________________________ Number of speech therapy sessions involved: ________________________

Did your laryngectomy involve:

1. Radical neck dissection? Yes ___ No ___ If yes, was it a complete or unilateral neck dissection?

2. Reconstruction of the pharynx? Yes ___ No ___

3. Reconstruction of the esophagus? Yes ___ No ___

4. Radiation therapy? Yes ___ No ___ If yes, preoperative or postoperative?
   How many kernels ___________ or Rads ___________?

5. Follow-up speech therapy? Yes ___ No ___ If yes, how much speech therapy did you receive and how successful was this therapy?

   ____________________________________________________________

   ____________________________________________________________

   ____________________________________________________________

   ____________________________________________________________

What is your primary method of producing speech at this time?

   ____________________________________________________________

   ____________________________________________________________

   ____________________________________________________________

   ____________________________________________________________

Please rate how good you are at speaking

Poor   Fair   Good   Excellent

How satisfied are you with your current method of speaking?

Very dissatisfied   Dissatisfied   Satisfied   Very Satisfied

Date of prosthesis fitting:

Number of speech therapy sessions used to train use of prosthesis:
If you could improve your voice, please indicate in what areas you would most like to have improvement, specify the type of improvement, and rank the areas in order of importance to you. For instance if you wanted to change only in the areas of loudness and pitch, and loudness was your first concern, you would put a 1 next to loudness and a 2 next to pitch and on the line following loudness you might put "Louder voice" or following pitch "Higher pitch."

____ Loudness
____ Pitch
____ Quality
____ Duration
____ Other

Do you feel that you have any hearing problems?  Yes  No

If yes, please describe your problem.

________________________

Do you have any vision problems that might interfere with your reading ability?  Yes  No  If yes, please describe.

________________________

If you are using a voice prosthesis, please describe any problems you may have had with it.

________________________
Background Information

Date_________________

Do not sign your name on the form. We want to preserve your right to privacy.

Participant __________________________ Site __________________________ Type ____________ No. __________________________

Age: _______ Sex: _______ Years of Education: __________________________

Occupation ____________________________

Do you smoke? __________. If "yes", for how long? __________________________

If "yes", how much? __________________________

Have you given up smoking? __________ How long ago? __________________________

How long did you smoke? __________________________

Have you ever had speech therapy? __________________________

If "yes", how long ago and for what? __________________________

__________________________________________

Do you have any hearing problems? __________________________

If "yes", please describe your problem. __________________________

__________________________________________

Do you have any vision problem that might interfere with your reading ability? __________________________

If "yes", please describe. __________________________
APPENDIX D

RAINBOW PASSAGE
WHEN THE SUNLIGHT STRIKES RAINDROPS IN THE AIR, THEY ACT LIKE A PRISM AND FORM A RAINBOW. THE RAINBOW IS A DIVISION OF WHITE LIGHT INTO MANY BEAUTIFUL COLORS. THESE TAKE THE SHAPE OF A LONG ROUND ARCH, WITH ITS PATH HIGH ABOVE AND ITS TWO ENDS APPARENTLY BEYOND THE HORIZON. THERE IS, ACCORDING TO LEGEND, A BOILING POT OF GOLD AT ONE END. PEOPLE LOOK, BUT NO ONE EVER FINDS IT. WHEN A MAN LOOKS FOR SOMETHING BEYOND HIS REACH, HIS FRIENDS SAY HE IS LOOKING FOR THE POT OF GOLD AT THE END OF THE RAINBOW.

(FAIRBANKS, 1960)
APPENDIX E

ELEVEN "H__D" SENTENCES
1. I will say heed.
2. I will say hid.
3. I will say head.
4. I will say had.
5. I will say hod.
6. I will say hud.
7. I will say who'd.
8. I will say hood.
9. I will say hoed.
10. I will say hawed.
11. I will say heard.
APPENDIX F

RECORDING SESSION PROCEDURES
Recording Session Procedures

Set-up

1. Turn on cassette deck and tone generator

2. Check deck switches: CrO₂, MIC, Dolby (Mic. for subjects, line for reference signal)

3. Attenuator

4. Connect headset to left channel

5. Insert tape

6. Push "Pause"

7. Push "Play/Record"

8. Test headset

9. Inventory forms: a. Permission sheet; b. Questionnaire; c. Study description; d. Rainbow Passage; e. 11 sentences; f. Appendix A

10. Record 4/sec.ref.sign (1000 Hz 84.5 dBSPL or 100 dBHL at 1/4") = 0 V.U.

Latency

Equipment: Recording form, pencil, stopwatch

Clinician's Actions: Assign ID#. Check 9a. & 9b. Read instructions to subject. RELEASE PAUSE. ID subject on tape. Start watch and signal for phonation. Stop watch at audible phonation. Record Time, Watch V.U.

Directions to Subject: When I say now, I want you to say /a/ as fast as you can. I want as little delay as possible. Only say /a/ once for each time I say now. We will do this 5 times. Are you ready?

Consistency

Equipment: Recording forms, pencil

Clinician's Actions: Read instructions to client. Criteria: One attempt is defined as a single occlusion of the stoma, single oral movement to phonate, or a single attempt to charge. Any attempt to stop and start over is scored as a failure in the minus column. Any attempt which is unsuccessful after 3 sec. is also scored as a failure. Audible phonation on a single attempt indicated a success and is scored in the plus column. Signal the subject to begin. Remember V.U.

Directions to Subject: When I say now, I want you to say /a/. There is no need to rush, simply take your time and say /a/ once when I say now. We
will try this measure 10 times. Do you have any questions? Are you ready?

**Duration**

**Equipment:** Recording form, pencil, stopwatch.

**Clinician's actions:** Read the instructions to the subject. Signal the subject when to begin. Start the watch as soon as audible phonation is achieved. Stop the watch as soon as phonation ceases. Record elapsed time. V.U.

**Directions to Subject:** This time we are looking at how long you can make a sound last. Just as in the two previous tasks, when I say now, I want you to say /a/ once. But this time, say /a/ for as long as you can on only one breath. We will have 5 trials this time. Do you have any questions? Ready?

**/da/ repetitions**

**Equipment:** Recording form, pencil

**Clinician's actions:** Read the instructions to the subject. Signal the subject when to begin. Count the number of syllables produced on one inspiration/expiration of air. Record V.U.

**Directions to Subject:** This time instead of measuring how long you can make a sound last, we will look at how many times you can say /da/ with just one breath. We will try this 5 times. When I say now, start saying /da/ as often as you can until you are out of breath, then stop. Do you have any questions? Are you ready?

**Oral Reading (May require attenuator)**

**Equipment:** Recording form, pencil, "Rainbow Passage", stopwatch, GLASSES(?)

**Clinician's actions:** Read the directions to the client. Emphasize that once time has started, the client is not to go back and repeat any sentences. If the client does go back and repeat a sentence, stop the client and the stopwatch and start the task over. Since the occasional repetition of a word may indicate some difficulty which the client may be experiencing in producing voice, these repetitions are allowed and will reflect poorer pharyngeal efficiency through increased time. STOP RECORDER. Once the silent reading is completed, start recorder and signal the client to begin. Start the stopwatch at the onset of phonation, stop the watch at the end of the paragraph or after three minutes, whichever comes first. Record the elapsed time.

**Directions to Subject:** I want you to read this paragraph to me, but first read it silently to yourself so that you are familiar with the words. If you have any questions, please ask me. After you have read this silently, when I say now, please read the paragraph to me at a comfortable rate. Once you start reading, do not back up and start over. Just read through the entire passage. Do you have any questions?
**Sentence reading** (may require attenuator)

Equipment: Same as above. Replace Passage with 11 sentences. Minus stopwatch.
Clinician's actions: read the instructions and the 11 sentences to the subject. Signal when to begin.
**Directions to Subject:** We have only 2 more exercises to go. The next thing I want you to do is to read these sentences at your normal rate and volume, but pause for about one second at the end of each sentence. Before you begin, I will read each sentence to you. Do you have any questions? Are you ready?

**Intensity Range** (may require attenuator)

Equipment: None
Clinician's actions: Read the directions to the subject. Emphasize the subject should produce only one phonation per signal. Signal the subject to begin.
**Directions to Subject:** This is the last exercise. Please say /a/ when I say now. First say /a/ as quietly as possible while still voicing and without whispering. We will do this 3 times. Only say /a/ once for each time I say now. Once we are done with the quiet /a/, we will do the same thing except that I want you to say /a/ as loudly as is comfortable for you. We will do these comfortably loud /a/'s 3 times also. Do you have any questions? Alright, first /a/ as quietly as you can and not whisper. Now /a/ three times as loudly as is comfortable for you.
APPENDIX G

VERBAL DIRECTIONS TO LISTENERS
To assist you in evaluating the voices on this tape, a scale of five voices has been collected. The voices on this scale represent varying degrees of acceptability, with the first voice representing the least acceptable example and the fifth voice representing the most acceptable example. Your job is to associate these representative examples with the experimental voice samples on this tape. As you hear each experimental sample decide which of the five representative voices the experimental sample most closely resembles and mark your form with the appropriate number and sex. To familiarize you with the five point scale, we will listen to the entire scale twice in order of increasing acceptability, then to a random order of the voices. During this random order, please try to identify which voice you are hearing.

—Scale presented twice—

Now we will listen to the representative voices in random order. Try to identify which number the voice is identified with.

—Voices 1-5 randomly presented twice—

Now to the main problem. You will hear 55 voice samples. Rate them on the 1-5 scale in terms of how acceptable each voice is and based upon how closely each voice resembles one of the examples which we have just listened to. This is a forced choice task, so please leave no blanks.

To assist you, after every fifth sample, a representative voice sample from the scale will be provided.

Once again here is the entire scale in order of least to most acceptable.

Thank you.
APPENDIX H

ACCEPTABILITY FORM FOR SCALE CONSTRUCTION
Acceptability Rating

Age:  Sex:  Native Language:  
Are you familiar with alaryngeal speech?

You are requested to evaluate the following speech samples with respect to how acceptable the quality of each sample is to you, and to determine whether the speaker is male or female. Evaluate the samples on a scale of 1-9 with 1, the lowest number, as least acceptable, and 9, the highest number, as most acceptable. Each sample will present the same two sentences with a 5 second pause following to allow you to mark your score sheet. The total number of samples which you are asked to evaluate is 50.

A scale of 1-9 is provided for each sample. Please circle the score for each sample.

In making your judgements about the speakers you are about to hear, give careful attention to the following attributes: pitch, rate, understandability, voice quality, loudness, and the amount of noise that may be present in the voice. Your basic consideration should be how pleasant or unpleasant each voice is for you to listen to.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Score</th>
<th>Sample</th>
<th>Score</th>
<th>Sample</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1 2 3 4 5 6 7 8 9</td>
<td>21</td>
<td>1 2 3 4 5 6 7 8 9</td>
<td>41</td>
<td>1 2 3 4 5 6 7 8 9</td>
</tr>
<tr>
<td>2</td>
<td>1 2 3 4 5 6 7 8 9</td>
<td>22</td>
<td>1 2 3 4 5 6 7 8 9</td>
<td>42</td>
<td>1 2 3 4 5 6 7 8 9</td>
</tr>
<tr>
<td>3</td>
<td>1 2 3 4 5 6 7 8 9</td>
<td>23</td>
<td>1 2 3 4 5 6 7 8 9</td>
<td>43</td>
<td>1 2 3 4 5 6 7 8 9</td>
</tr>
<tr>
<td>4</td>
<td>1 2 3 4 5 6 7 8 9</td>
<td>24</td>
<td>1 2 3 4 5 6 7 8 9</td>
<td>44</td>
<td>1 2 3 4 5 6 7 8 9</td>
</tr>
<tr>
<td>5</td>
<td>1 2 3 4 5 6 7 8 9</td>
<td>25</td>
<td>1 2 3 4 5 6 7 8 9</td>
<td>45</td>
<td>1 2 3 4 5 6 7 8 9</td>
</tr>
<tr>
<td>6</td>
<td>1 2 3 4 5 6 7 8 9</td>
<td>26</td>
<td>1 2 3 4 5 6 7 8 9</td>
<td>46</td>
<td>1 2 3 4 5 6 7 8 9</td>
</tr>
<tr>
<td>7</td>
<td>1 2 3 4 5 6 7 8 9</td>
<td>27</td>
<td>1 2 3 4 5 6 7 8 9</td>
<td>47</td>
<td>1 2 3 4 5 6 7 8 9</td>
</tr>
<tr>
<td>8</td>
<td>1 2 3 4 5 6 7 8 9</td>
<td>28</td>
<td>1 2 3 4 5 6 7 8 9</td>
<td>48</td>
<td>1 2 3 4 5 6 7 8 9</td>
</tr>
<tr>
<td>9</td>
<td>1 2 3 4 5 6 7 8 9</td>
<td>29</td>
<td>1 2 3 4 5 6 7 8 9</td>
<td>49</td>
<td>1 2 3 4 5 6 7 8 9</td>
</tr>
<tr>
<td>10</td>
<td>1 2 3 4 5 6 7 8 9</td>
<td>30</td>
<td>1 2 3 4 5 6 7 8 9</td>
<td>50</td>
<td>1 2 3 4 5 6 7 8 9</td>
</tr>
<tr>
<td>11</td>
<td>1 2 3 4 5 6 7 8 9</td>
<td>31</td>
<td>1 2 3 4 5 6 7 8 9</td>
<td>51</td>
<td>1 2 3 4 5 6 7 8 9</td>
</tr>
<tr>
<td>12</td>
<td>1 2 3 4 5 6 7 8 9</td>
<td>32</td>
<td>1 2 3 4 5 6 7 8 9</td>
<td>52</td>
<td>1 2 3 4 5 6 7 8 9</td>
</tr>
<tr>
<td>13</td>
<td>1 2 3 4 5 6 7 8 9</td>
<td>33</td>
<td>1 2 3 4 5 6 7 8 9</td>
<td>53</td>
<td>1 2 3 4 5 6 7 8 9</td>
</tr>
<tr>
<td>14</td>
<td>1 2 3 4 5 6 7 8 9</td>
<td>34</td>
<td>1 2 3 4 5 6 7 8 9</td>
<td>54</td>
<td>1 2 3 4 5 6 7 8 9</td>
</tr>
<tr>
<td>15</td>
<td>1 2 3 4 5 6 7 8 9</td>
<td>35</td>
<td>1 2 3 4 5 6 7 8 9</td>
<td>55</td>
<td>1 2 3 4 5 6 7 8 9</td>
</tr>
<tr>
<td>16</td>
<td>1 2 3 4 5 6 7 8 9</td>
<td>36</td>
<td>1 2 3 4 5 6 7 8 9</td>
<td>56</td>
<td>1 2 3 4 5 6 7 8 9</td>
</tr>
<tr>
<td>17</td>
<td>1 2 3 4 5 6 7 8 9</td>
<td>37</td>
<td>1 2 3 4 5 6 7 8 9</td>
<td>57</td>
<td>1 2 3 4 5 6 7 8 9</td>
</tr>
<tr>
<td>18</td>
<td>1 2 3 4 5 6 7 8 9</td>
<td>38</td>
<td>1 2 3 4 5 6 7 8 9</td>
<td>58</td>
<td>1 2 3 4 5 6 7 8 9</td>
</tr>
<tr>
<td>19</td>
<td>1 2 3 4 5 6 7 8 9</td>
<td>39</td>
<td>1 2 3 4 5 6 7 8 9</td>
<td>59</td>
<td>1 2 3 4 5 6 7 8 9</td>
</tr>
<tr>
<td>20</td>
<td>1 2 3 4 5 6 7 8 9</td>
<td>40</td>
<td>1 2 3 4 5 6 7 8 9</td>
<td>60</td>
<td>1 2 3 4 5 6 7 8 9</td>
</tr>
</tbody>
</table>
APPENDIX I

ACCEPTABILITY FORM FOR
ASSESSING EXPERIMENTAL VOICES
Age: Sex: Native Language:
Are you familiar with alaryngeal speech?

You are requested to evaluate the following speech samples with respect to how acceptable the quality of each sample is to you, and to determine whether the speaker is male or female. Evaluate the samples on a scale of 1-5 with 1, the lowest number, as least acceptable, and 5, the highest number, as most acceptable. Each sample will present the same two sentences with a 2 second pause following to allow you to mark your score sheet. The total number of samples which you are asked to evaluate is 55.

A scale of 1-5 is provided for each sample. Please circle the score for each sample. In the blank to the right of each number, put an M if you think the subject is male and F if you think the subject is female.

In making your judgements about the speakers you are about to hear, give careful attention to the following attributes: pitch, rate, understandability, voice quality, loudness, and the amount of noise that may be present in the voice. Your basic consideration should be how pleasant or unpleasant each voice is for you to listen to.

Remember this is an assessment of voice quality, not of reading ability. Some subjects are poorer readers than others, please do not allow reading ability to affect your judgements. Try to judge each voice as if you were listening to that subject in a conversation, not to a reading passage.

Sample A-J are practice samples.

<table>
<thead>
<tr>
<th>Sex</th>
<th>Acceptability</th>
<th>Sex</th>
<th>Acceptability</th>
<th>Sex</th>
<th>Acceptability</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. 1</td>
<td>2 3 4 5</td>
<td>13.</td>
<td>1 2 3 4 5</td>
<td>35.</td>
<td>1 2 3 4 5</td>
</tr>
<tr>
<td>B. 1</td>
<td>2 3 4 5</td>
<td>14.</td>
<td>1 2 3 4 5</td>
<td>36.</td>
<td>1 2 3 4 5</td>
</tr>
<tr>
<td>C. 1</td>
<td>2 3 4 5</td>
<td>15.</td>
<td>1 2 3 4 5</td>
<td>37.</td>
<td>1 2 3 4 5</td>
</tr>
<tr>
<td>D. 1</td>
<td>2 3 4 5</td>
<td>16.</td>
<td>1 2 3 4 5</td>
<td>38.</td>
<td>1 2 3 4 5</td>
</tr>
<tr>
<td>E. 1</td>
<td>2 3 4 5</td>
<td>17.</td>
<td>1 2 3 4 5</td>
<td>39.</td>
<td>1 2 3 4 5</td>
</tr>
<tr>
<td>F. 1</td>
<td>2 3 4 5</td>
<td>18.</td>
<td>1 2 3 4 5</td>
<td>40.</td>
<td>1 2 3 4 5</td>
</tr>
<tr>
<td>G. 1</td>
<td>2 3 4 5</td>
<td>19.</td>
<td>1 2 3 4 5</td>
<td>41.</td>
<td>1 2 3 4 5</td>
</tr>
<tr>
<td>H. 1</td>
<td>2 3 4 5</td>
<td>20.</td>
<td>1 2 3 4 5</td>
<td>42.</td>
<td>1 2 3 4 5</td>
</tr>
<tr>
<td>I. 1</td>
<td>2 3 4 5</td>
<td>21.</td>
<td>1 2 3 4 5</td>
<td>43.</td>
<td>1 2 3 4 5</td>
</tr>
<tr>
<td>J. 1</td>
<td>2 3 4 5</td>
<td>22.</td>
<td>1 2 3 4 5</td>
<td>44.</td>
<td>1 2 3 4 5</td>
</tr>
<tr>
<td>1.</td>
<td>2 3 4 5</td>
<td>23.</td>
<td>1 2 3 4 5</td>
<td>45.</td>
<td>1 2 3 4 5</td>
</tr>
<tr>
<td>2.</td>
<td>2 3 4 5</td>
<td>24.</td>
<td>1 2 3 4 5</td>
<td>46.</td>
<td>1 2 3 4 5</td>
</tr>
<tr>
<td>3.</td>
<td>2 3 4 5</td>
<td>25.</td>
<td>1 2 3 4 5</td>
<td>47.</td>
<td>1 2 3 4 5</td>
</tr>
<tr>
<td>4.</td>
<td>2 3 4 5</td>
<td>26.</td>
<td>1 2 3 4 5</td>
<td>48.</td>
<td>1 2 3 4 5</td>
</tr>
<tr>
<td>5.</td>
<td>2 3 4 5</td>
<td>27.</td>
<td>1 2 3 4 5</td>
<td>49.</td>
<td>1 2 3 4 5</td>
</tr>
<tr>
<td>6.</td>
<td>2 3 4 5</td>
<td>28.</td>
<td>1 2 3 4 5</td>
<td>50.</td>
<td>1 2 3 4 5</td>
</tr>
<tr>
<td>7.</td>
<td>2 3 4 5</td>
<td>29.</td>
<td>1 2 3 4 5</td>
<td>51.</td>
<td>1 2 3 4 5</td>
</tr>
<tr>
<td>8.</td>
<td>2 3 4 5</td>
<td>30.</td>
<td>1 2 3 4 5</td>
<td>52.</td>
<td>1 2 3 4 5</td>
</tr>
<tr>
<td>9.</td>
<td>2 3 4 5</td>
<td>31.</td>
<td>1 2 3 4 5</td>
<td>53.</td>
<td>1 2 3 4 5</td>
</tr>
<tr>
<td>10.</td>
<td>2 3 4 5</td>
<td>32.</td>
<td>1 2 3 4 5</td>
<td>54.</td>
<td>1 2 3 4 5</td>
</tr>
<tr>
<td>11.</td>
<td>2 3 4 5</td>
<td>33.</td>
<td>1 2 3 4 5</td>
<td>55.</td>
<td>1 2 3 4 5</td>
</tr>
<tr>
<td>12.</td>
<td>2 3 4 5</td>
<td>34.</td>
<td>1 2 3 4 5</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>