A Comparison of Esophageal and Electrolarynx Speech Using Audio and Audiovisual Recordings

An Honors Thesis (ID 499)

Mary L. Salzmann

Dr. William Kramer

Ball State University

Muncie, Indiana

May 1987

Spring Quarter 1987
TABLE OF CONTENTS

I. Introduction ........................................... 1

II. Review of Related Literature ......................... 2
    Esophageal speech .................................... 2
    Electrolarynx speech ................................. 3
    Research ............................................... 5
    Quality of alaryngeal speech ....................... 9

III. Method .................................................. 12
    Subjects ............................................... 12
    Instrumentation ....................................... 12
    Procedures ............................................ 12
    Statistical design and analysis of data .......... 13

IV. Results ................................................ 14
    Charts ............................................... 16

V. Discussion ............................................. 17

VI. References ............................................. 18
INTRODUCTION

There are several alaryngeal speech methods presently available to the laryngectomized population—esophageal speech, the electrolarynx, and the tracheoesophageal puncture, for example. These methods are used with varying degrees of satisfaction by the patients. To the average college-age student who has had no previous contact with laryngectomees, however, the speech methods may be met with a wide range of reactions and preferences. The purpose of this study is to survey the listener satisfaction of a group of college-age students who have had no previous contact with alaryngeal speech to determine their reactions and ratings of esophageal speech as compared to electrolarynx speech. Comparisons will also be made between the data obtained from audiovisual presentation and from audio presentation only.

Several studies have been performed to collect data comparing the acceptability of specific alaryngeal speech methods. Of these, many of the subjects chosen to rate the alaryngeal speaker were speech-language pathologists. Clinicians in the field of speech-language pathology may already have strong preferences for a particular alaryngeal speech method which will influence their ratings. This study will concentrate on the reactions and ratings of a group of college students who have never encountered esophageal speech or electrolarynx speech, neither visually or aurally.
REVIEW OF THE LITERATURE

The removal of a larynx leaves the patient without any form of speech communication, necessitating the assistance of a team of professionals to help him adjust to his new condition. The primary goal of the speech-language pathologist is to work with the laryngectomee to help him develop functional speech by the use of an alaryngeal speech method. The term "alaryngeal speech", first introduced by Kallen and later popularized by Diederich and Youngstrom, refers to any type of phonation used without a larynx, such as esophageal voice, buccal speech, and the use of electronic larynges (Boone, 1977). Esophageal voice and the use of the electronic larynx are the two alaryngeal speech methods focused on in this study.

Esophageal speech is produced by taking air into the esophagus and producing sound by releasing air through the vibration of the upper esophageal tract (Aronson, 1980). Three methods of air intake may be used--injection, inhalation, and swallowing. The method with typically produces the best sound is a combination of injection and inhalation. Effective esophageal speech is characterized by the following:

1. Reliable phonation on demand.
2. Rapid air intake.
3. Short latency between air intake and phonation.
4. Four to nine syllables per air charge.
5. Two to three seconds duration per air intake.
6. 85-129 words per minute.
7. Fundamental frequency of 52-82 Hz.
8. An average intensity of 6-7 dB below normal.
9. Good intelligibility.

(Aronson, 1980).

Not all laryngectomees are capable of developing effective esophageal speech due to physical or psychological factors. According to Salmon (1979) evidence supports that at least 1/3 of the laryngectomized population fail to achieve esophageal communication. There are a number of laryngectomees using functional esophageal speech; literature claims success rates of between 25% and 98% (Schaefer and Johns, 1982). Duguay cites that the literature ranges from a "low of 30% poor results to a high of 57.4% failure rate." (Schaefer and Johns, 1982). Traditionally, rehabilitation efforts have centered on the acquisition of esophageal speech. More recently, there has been a change in attitude and the use of the electrolarynx has been employed as an aid in esophageal speech acquisition programs.

Today the electronic artificial larynx is frequently used to provide effective immediate voice production. This prosthesis is more widely accepted by speech pathologists due to the increasing emphasis on meeting the patients' communication needs rather than a particular speech mode such as esophageal speech (Weiss, 1985). The first electrolarynx was developed by the Bell Telephone
Laboratories in 1959 and was marketed through the Western Electric Company. The most recent version of this unit, developed in 1985, is the Model C (Blom and Salmon, 1986). There are several brands of electronic larynges: 1.) The Western Electric Electronic Larynx Model 5C is a neck-type, battery powered and sells for about $100. 2.) The Aurex Neovox Electronic Larynx by the Aurex Corporation, is a neck-type, battery powered (rechargeable), and sells for $250-300. 3.) The Siemens Servox Electronic Larynx, sold by Siemens Corporation, is a neck-type, battery powered model which sells for $400-500. 4.) The Barts Vibrator Electronic Larynx sold by Park Surgical Company is a neck-type, battery powered model which sells for $200-250. 5.) The Cooper-Rand Electronic Larynx sold by Luminaud for $150-200 is a tube-in-mouth type, battery powered model (Prater and Swift, 1984). The most widely used type of electrolarynx is a hand-held model which is placed firmly on the neck on the area which is the best for the individual patient. The model has a vibrating diaphragm which transmits the vibrations through the skin of the neck and into the hypopharynx (Prater and Swift, 1984).

The selection of the alaryngeal speech method to be used depends on the following factors (Prater and Swift, 1984): 1.) physical factors such as the degree of tissue loss, 2.) the noise level of the environment in which the laryngectomized patient will have to communicate, 3.) the patients' level of motivation for learning an alaryngeal
speech method, and 4.) the personal preference of the patient. According to Boone (1977), it is frequently stated that "if the new laryngectomee uses an artificial larynx before learning esophageal speech, he will never develop esophageal speech." This has been disputed by several authors (Boone, 1977; Prater and Swift, 1984; and Salmon and Goldstein, 1978); and, as Duguay states, "I have never heard of any case or seen a published report indicating that the use of an artificial larynx negated the acquisition of esophageal speech" (Salmon and Goldstein, 1978). It has been proven that the use of the electrolarynx has been beneficial in the training of esophageal speech.

There are several advantages and disadvantages of esophageal speech and electrolarynx speech. Besides the advantage of providing immediate effective speech while in an esophageal training program, the following are other benefits of artificial larynges: 1.) provides immediate speech after laryngeal surgery, 2.) provides speech method for laryngectomees incapable of producing functional esophageal speech, 3.) takes less time to learn, 4.) provides a higher intensity level than esophageal speech, and 5.) provides a temporary alternative to esophageal speech when fatigued or upset (Prater and Swift, 1984). Some of the possible disadvantages of artificial larynges, according to Salmon and Goldstein (1978), are that it produces an inhuman sound quality, is a sign of defeat, is used as a crutch, requires the use of a hand, is expensive
to maintain, is bulky, calls attention to lack of the ability to speak normally, and makes casual remarks awkward to produce.

As in artificial larynx speech, there are several advantages and disadvantages to the esophageal speech method. There are various undesirable behaviors of esophageal speech; these include unnecessary stoma noise, excessive swallowing of air, distracting noise, and facial grimaces (Boone, 1977). Esophageal speech requires more training time, is not as effective in noisy environments, and involves latency between air intake and phonation which causes pauses in speech. The advantages of esophageal speech are closely related to the disadvantages of electronic laryngeal speech—it does not require an instrument or the use of either hand, sounds more natural, and does not cost anything.

Many studies have investigated the preferences between esophageal and electronic laryngeal speech revealing contradictory results. In an early study conducted by Hyman in 1955, college students judged audiotape recordings of esophageal speakers, artificial larynx speakers, and normal speakers. The data revealed that artificial larynx speakers were preferred in comparison to esophageal speakers (Salmon and Goldstein, 1978). In later studies (Crouse, 1962; Shames, Font, and Matthews, 1963; Bennett and Weinberg, 1973; and Kalb, 1977), esophageal speakers were preferred in
comparison to artificial laryngeal speakers (Salmon and Goldstein, 1978). The conditions of the latter studies varied. In the Crouse study audiovisual presentation was used and the two groups of speakers were rated by both a group of trained speech pathologists and a group of trained speech pathologists and a group of nonprofessionals; both groups preferred esophageal speech over artificial laryngeal speech. Esophageal speakers were rated superior to artificial laryngeal speakers on one word intelligibility in the Shames, Font, and Matthews study; however, no significant difference was noted in sentence intelligibility. In 1973, Bennett and Weinberg compared the acceptability ratings of esophageal, artificial laryngeal speech, and normal speech, employing proficient laryngectomees and a group of naive listeners. In this study, which used audiotape samples and the reading of a standard passage, the results showed a preference of esophageal over artificial laryngeal speech. According to Salmon and Goldstein (1978), in 1977 Kalb compared the intelligibility of esophageal and artificial laryngeal speech produced by different speakers as well as by the same speaker, producing interesting results. Using 30 naive listeners to determine word intelligibility, Kalb discovered that there was no significant difference between esophageal speech and artificial laryngeal speech when each method was produced by the same speaker; however, when the two speech methods were used by different speakers, esophageal speakers
were rated more intelligible than artificial laryngeal speakers (Salmon and Goldstein, 1978). The results of this study may have been influenced by the various levels of proficiency of the individual speakers.

A study performed in 1963 by McCroskey and Mulligan may indicate perceptual differences between trained professional listeners and naive listeners. Esophageal speech was rated more intelligible than artificial laryngeal speech by professional and student listeners; in contrast, naive listeners indicated a higher intelligibility rating for the artificial larynx speakers than the esophageal speakers. Esophageal speech was ranked by 60% of the listeners as the least pleasant of three methods of alaryngeal speech in a study of message-to-competition ratios (Clark and Stemple, 1982). Data from the Green and Hults study (1982) indicated that of the following speech methods—Tokyo pneumatic aid, the Servox electrolarynx, poor esophageal speech, and normal laryngeal speech—poor esophageal speech was the least preferred on voice quality, pitch, loudness, intelligibility, and rate by naive judges. As indicated by the previously stated results, a review of the literature on comparisons between esophageal speech and artificial larynx speech indicates differences in preferences between these two methods. As stated by Clark and Stemple, these differences may be attributed to "differences among studies relating to differing speaking proficiencies of the alaryngeal talkers, differing listening environments and/or
conditions, and different perceptions of the speech samples as the result of differing preference-judgment instructions provided (Clark and Stemple, 1982).

The laryngectomee's perception of his speech is often influenced by the listener's impression in the social situation. When a laryngectomized individual experiences success with his new voice with a listener, he is encouraged to continue in other environments. As Duguay states, "if he tries and cannot be understood, he is apt to experience frustration and a consequent diminished desire to talk" (Duguay, 1986). Success is achieved by the use of the appropriate speech aid for the individual and encouragement on the part of the clinician and his significant others to practice speech in different environments, such as in a restaurant or in a grocery store. Practicing esophageal and/or electrolarynx speech in various social situations or settings requires the laryngectomee to adapt his speech to the environment.

The laryngectomee must make adjustments for effective communication due to differences between the dimensions of his new voice and that of a laryngeal speaker. Contrasts have been noted in intensity, quality, rate, intelligibility, and frequency by several researchers. The average intensity level of esophageal voice is 40-50 dB, which is up to about 20-25 dB lower than laryngeal voice (Perry, 1983). The esophageal intensity range is only about
20dB, compared to 45 dB in the normal speaker. Intensity is related to intelligibility; "The apparent loudness of a voice is also a function of the speaker's intelligibility" (Duguay, 1986). It has also been noted that the reduction of intensity improves the quality of the alaryngeal voice (Shanks, 1986). Some vocal qualities attributed to esophageal voice are rough, hoarse, strain, tense, wet, and bubbly (Weinberg, 1986). Weinberg and Bennett (1973) state that the most frequent complaint of listeners of esophageal speech is abnormal quality (Hyman, 1986). Another difference among voice attributes of the laryngeal and alaryngeal speaker is in the rate of speech. The average adult reads at a rate of 150-165 words per minute, whereas the average good esophageal speaker reads at a rate of 80-130 words per minute; this translates to the assumption that in a normal speaking situation, esophageal speakers have a lower rate. This has not yet been significantly correlated with acceptability or intelligibility of the laryngectomized population (Salmon, 1983). As stated earlier, it has been suggested to decrease intensity in order to improve the intelligibility of alaryngeal speech. In the area of frequency, for the average male, the pitch of esophageal voice is one octave lower and for the average female, one to two octaves lower (Shanks, 1986). Although the frequency of alaryngeal speech is typically lower, this has not been correlated with intelligibility in either alaryngeal or laryngeal speakers (Hyman, 1986).
These differences in attributes and other factors contribute to the listener's perception of alaryngeal speech. In a social interaction study conducted by Blood and Blood, it was conclude that laryngeal individuals prefer to interact with laryngectomees who acknowledge their handicap rather than those who do not. The subjects perceived the acknowledging individual to be more pleasant, calm, active, likable, well-adjusted, stronger, tougher, and hardworking on a bipolar adjective scale (Blood and Blood, 1982). In 1971, Hoops and Noll studied the effect of visual aspects of esophageal speech on listener perceptions and conclude that judgments of esophageal speech based on audiovisual presentation differed significantly from ratings based on auditory or visual information only (Hubbard and Kushner, 1980). The laryngectomee's fear of being rejected by others in social situations can hamper his learning of alaryngeal speech. In a study conducted by Diedrich and Youngstrom 50% of 72 patients who responded stated that there were continuing social situations where they suffered intense embarrassment over their speech. Some of the factors which help laryngectomees overcome these anxieties are their sheer determination to speak, their conclusions that they have no other choice, and their resignation to "make the best of it" (Diedrich, 1966).
METHODOLOGY

Subjects

Sixty naive college-age listeners will be selected as subjects for the study. These students will be randomly selected from a class (or classes) at Ball State University. The students will have not had previous social contact with laryngectomized individuals. Two speakers will be used—and excellent esophageal speaker and an excellent electrolarynx speaker.

Instrumentation

The proficiency of the two speakers will be judged by a highly qualified speech-language professional. Two methods of presentation will be used—audiotape and videotape. The students will rate the esophageal and electrolarynx speakers using the Osgood Semantic Differential. Using the set of bipolar adjectives, they will rate each speaker separately using a seven-point scale. Each presentation will consist of the speaker counting from 1 to 20, saying the months of the year, and reciting a short passage.

Procedure

Thirty of the randomly selected students will be presented with a 2-minute audiotape presentation of the electrolarynx speaker and the esophageal speaker. The other thirty students will be presented with a 2-minute videotape
The presentation of the electrolarynx speaker and the esophageal speaker. The split half technique will be used to control the sequence and order effect in the experiment. For example, for the audiovisual presentation, 15 of the subjects will view the electrolarynx speaker first; the other 15 subjects will view the esophageal speaker first.

Statistical Design and Analysis of Data

The statistical design to be employed will be the Anova 2x2 Test of Variance. This design will compare two sets of variables—the method of speech (electrolarynx and esophageal) and the method of presentation (audiovisual and audio only) as rated by naive college student listeners using the Osgood Semantic Differential task.
RESULTS

Due to limited time available, the method of presentation of the audiotape and audiovisual tapes was altered slightly. Both tapes of each speaker were shown to two groups of students. The order of presentation was varied using the "split-half" technique to control for sequence and order effects.

The 2-Way Analysis of Variance results showed an initial F ratio of 12.55, which is significant at the .01 level. Subsequent investigation showed the primary source of variation to be the type of alaryngeal speech used, not the modality of recording (audio vs. visual). The alpha for the degrees of freedom of (1, 63) was 7.04 at the .01 level. Results show an F ratio of 33.97, which is significant at this level.

Overall listener satisfaction was computed in percentages. The following are the calculated means of each group: (1.) Audio only, esophageal speaker--69, (2.) Audio only, electrolarynx speaker--79, (3.) Audiovisual, esophageal speaker--70, (4.) Audiovisual, electrolarynx speaker--85.

A t-test was used to compare the type of alaryngeal speech used. On a t-test comparing the two types presented audiovisually, the t-test statistic was 4.56. This is
significant at the .01 level (alpha=2.701) with the degrees of freedom (1, 42). This shows a significant difference between the two types of speech using the audiovisual modality of recording.

A t-test was also used to compare the esophageal and electrolarynx speakers within the audiotape only presentation. A t-test statistic of 3.50 was calculated. This is significant at the .01 level (alpha=2.701) with the degrees of freedom (1, 42).
### CHARTS

<table>
<thead>
<tr>
<th>Source</th>
<th>d.F.</th>
<th>S.S.</th>
<th>M.S.</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Method</td>
<td>3</td>
<td>3685.48</td>
<td>1228.496</td>
<td>12.547</td>
</tr>
<tr>
<td>Listener</td>
<td>21</td>
<td>2307.15</td>
<td>109.86</td>
<td></td>
</tr>
<tr>
<td>Error</td>
<td>63</td>
<td>6168.26</td>
<td>97.908</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>87</td>
<td>12160.89</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

alpha for d.F. of (3, 63) = 2.75 (.05 level)

4.11 (.01 level)

<table>
<thead>
<tr>
<th>Source</th>
<th>d.F.</th>
<th>S.S.</th>
<th>M.S.</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Audio vs. Visual</td>
<td>1</td>
<td>219.55</td>
<td>219.55</td>
<td>2.24</td>
</tr>
<tr>
<td>B</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Electro vs. Es</td>
<td>1</td>
<td>3325.91</td>
<td>3325.91</td>
<td>33.97</td>
</tr>
<tr>
<td>Interaction</td>
<td>1</td>
<td>140.01</td>
<td>140.01</td>
<td>1.43</td>
</tr>
<tr>
<td>Error</td>
<td>63</td>
<td>6168.26</td>
<td>97.908</td>
<td></td>
</tr>
</tbody>
</table>

alpha for d.F of (1, 63) = 3.99 (.05 level)

7.04 (.01 level)
CONSENT FORM

You are invited to participate in a research project concerning speech techniques used by laryngectomees, individuals who have had their larynges removed due to the effects of cancer. This project investigates college students' reactions to various types of alaryngeal speech (speech without use of the larynx).

Your participation in this project will require you to listen to an audiotape and/or to view an audiovisual tape of two laryngectomees. The tapes are less than five minutes in length. Following the presentation, you will complete a questionnaire asking you to rate the speaker(s) according to several attributes. The presentation and questionnaire will take you approximately fifteen minutes of your time.

You do not need to write your name on the questionnaire. All data collected in this study will be kept anonymous. There are no risks or ill effects from participating in this project. You are free to withdraw from participation at any time. You are also welcome to ask any questions of the investigator before signing the Consent Form and participating in the study.

INFORMED CONSENT STATEMENT

I, ____________________, agree to participate in this research project on alaryngeal communication. This study was clearly explained to me and any questions were answered to my satisfaction.

---------------------------------
Participant's Signature
DIRECTIONS

Thank you for participating in this research project. You will be guided through these simple instructions for your participation.

For each laryngectomee that you see and/or hear on tape you will be asked to fill out a questionaire. The questionaire consists of thirty items. Each item is a pair of attributes that each person may possess to a certain degree. The pairs are separated by a continuum of seven blanks. Think of these seven blanks as a continuous line from one end of the extreme to the other. You are asked to place a check mark on the blank from 1 to 7 which most closely matches your impression of this laryngectomee.

For each pair of items you should have a check mark somewhere between blank 1 and blank 7. Please complete the entire questionnaire. Respond with your first instincts, as they usually most accurately reflect your impression of the speaker.

Your help on this project is greatly appreciated. If you have any questions, please feel free to ask the investigator.
DISCUSSION

With this group of naive listeners, the individual using electrolarynx speech was viewed with an overall higher level of satisfaction than the esophageal speaker. This is supported by the results of the 2-Way Analysis of Variance, t-tests, and the calculated means of each group. There may be several reasons for this; age, general appearance, loudness level, microphone placement, observer field of vision, and tape recorder placement (listener). No significant difference was found between the two modalities of recording (audio only vs. audiovisual). With these two speakers, the primary source of variation was directly related to the actual speakers. Calculations indicate this variation may be due to the type of speech used.

In order to limit speaker variations, it is suggested that one speaker be used who is proficient in both types of alaryngeal speech. This would eliminate personal differences between speakers which may be evident from audio and/or audiovisual recordings. An increase in the number of listeners may also add credibility to these findings. Research in this area using a more diverse age group of listeners may give an accurate representation of the average social impression of the general public. This may be significant information for alaryngeal speakers, speech pathologists, and professionals in medicine and other related fields.


