ACOUSTIC PROPERTIES OF ARTIFICIAL LARYNX SPEECH

BETHANN MOFFET

Elk Valley Home Health Care, Fayetteville, TN

REBEKAH H. PINDZOLA

Auburn University, Auburn, AL

Information about the physical properties of various types of alaryngeal speech is important both to laryngectomized patients and speech-language pathologists providing rehabilitative services. Acoustical properties presumably influence, or underlie, listeners’ perceptions (Lehiste, 1976; Moffet, 1987; Rothman, 1978). Knowledge of objective, measurable differences among artificial larynx devices therefore is clinically important.

Over the years, acoustic and perceptual research have compared normal speech, "artificial larynx speech" from a variety of devices, esophageal speech, and tracheoesophageal speech of laryngectomized patients (Weinberg & Riekena, 1973; Goldstein, 1978; Rothman, 1978, 1982; Weiss, Yeni-Komshian, & Heinz, 1979; Stalker, Hawk, & Smaldino, 1982; Robbins, Fisher, & Blom, 1984). The majority of such comparative research has been perceptual in nature. With regard to our understanding of acoustics, some research suggests that artificial larynx speech is similar to other alaryngeal speech forms on some acoustic parameters. Other studies show artificial larynx to be significantly different—either better or worse than other alaryngeal speech forms under scrutiny. No doubt acoustic differences also exist between the various artificial devices, yet this is not clear in the literature as the type of device researched often is not specified or the brands are pooled together and differences obscured.

More research is needed which better defines the types of devices measured. Newer artificial device models, such as the AT&T 5 C, and those in current use require study. Information of this type will allow speech-language pathologists to provide laryngectomies with more information and freedom for choosing a postsurgical speech method.

It is estimated that only 30–60% of laryngectomies develop effective traditional esophageal speech (Putney, 1958; Gardner & Harris, 1981; Martin, 1963; King et al., 1968). These reports suggest that many laryngectomies will use artificial devices "as a last resort" to be able to talk. Fortunately, the newer tracheoesophageal puncture techniques and prostheses are enabling more laryngectomies to learn esophageal voice production. Success rates may be as high as 88% (Blom, Singer, & Hamaker, 1982). Still, according to Kelly (1983), acceptance of the artificial larynx as a primary source of postlaryngectomized speech must be increased in both social
and occupational environments or many laryngectomees will remain speechless or perceive themselves as failures.

Is artificial larynx speech so different as to render it useful only as a temporary method or only as a last resort if other speech rehabilitation attempts fail? Or, are the characteristics of artificial larynx speech such that it should be considered a primary and legitimate speech option? Perceptual studies (Bennett & Weinberg, 1973; Goldstein, 1978) indicate that “artificial larynx speech,” indeed, is acceptable. However, more research is needed on acoustical and perceptual aspects of modern, currently used types of artificial larynges in order to identify the sound characteristics and disadvantages or advantages of each device.

The present investigation collected acoustical data from alaryngeal speakers using artificial devices. Comparisons were made among two intraoral devices (Cooper-Rand and orally adapted Servox) and two neck placement devices (AT&T 5 C and standard Servox).

METHODS

Subjects

Subjects were identified by contacting regional laryngectomee clubs and rehabilitation hospitals. Fourteen male esophageal speakers were matched for age (within 8 years) and race (to control for dialect). All subjects were native speakers of American English living in Alabama or Georgia. Table 1 summarizes background information for each subject including age, level of education completed, amount of time since their laryngectomy, type of artificial larynx device owned (if any), and duration of past alaryngeal speech treatment. Although it was expected that alaryngeal speakers would be familiar with artificial larynges, persons who spoke
primarily with esophageal speech were selected as subjects. This was done to control for familiarity with a particular device and minimize its “practice effect.” All subjects read and signed an informed consent form prior to beginning the investigation which was approved by the Institutional Review Board of Auburn University.

SPEECH STIMULI AND PROCEDURES

Each volunteer subject was instructed in the use of each of the four artificial devices (Cooper-Rand, Servox, Intraoral Servox, and AT&T 5-C) according to Duguay’s (1983) training suggestions, and given a limited amount of time to practice with each device. Then each subject was taped on a Tascam 22-2 Reel-to-Reel recorder in a quiet room with a mouth-to-microphone distance of 3 inches. Alaryngeal subjects were recorded using each of 4 artificial devices; the order of device use was counterbalanced among subjects to randomize effects of practice.

All subjects read aloud a 98-word phonetically represented paragraph from the “Rainbow Passage” with each of the artificial larynges. Following instructions and a demonstration, the vowel /i/ was prolonged using each device. The vowel /i/ was chosen for its consistency in displaying jitter (Wilcox & Hori, 1980) and so was later analyzed for pitch perturbation.

ANALYSES

This investigation concerned measurable differences between artificial larynx devices of the intraoral and neck types. Five acoustic measures were made: fundamental frequency, fundamental frequency range, relative intensity, jitter, and rate of speech.

**Fundamental frequency.** The second sentence of the “Rainbow Passage” was analyzed to find the average fundamental frequency of each speech sample. A Visi-Pitch/Apple II E computer interface (Kay Elemetrics) with statistical subroutines was used to obtain the frequency analyses. With this procedure, the average fundamental frequency is calculated using the number of pulses (between the cursors marking the sentence) divided by time (duration between the same cursors).

**Fundamental frequency range.** Intonation was measured as fundamental frequency range within the arbitrarily located second sentence of the “Rainbow Passage,” as portrayed by cursor placement, using the same Visi-Pitch system. Fundamental frequency range is displayed as the highest (maximum Hz) and the lowest (minimal Hz) difference. The frequency output of each artificial larynx was preset to standardize investigation of this inflectional variability by the user.

**Relative intensity.** Although not a true measure of power, average intensity of the second sentence of the “Rainbow Passage” was obtained with the Visi-Pitch system (internal referent = 30 dB). The relative nature of this measure allows for only comparative statements regarding loudness of each device some 3 inches from the sound source.

**Jitter/Pitch perturbation.** Jitter, the cycle-to-cycle variation in vibratory frequency, was measured during prolongation of the vowel /i/. Again, the Visi-Pitch system was used for this analysis.
TABLE 2. Composite results of ANOVA procedures on acoustic measures.

<table>
<thead>
<tr>
<th>ANOVA Test</th>
<th>Factor*</th>
<th>F Value</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fundamental Frequency</td>
<td>A</td>
<td>0.17</td>
<td>none</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>5.87</td>
<td>significant at .05</td>
</tr>
<tr>
<td></td>
<td>A×B</td>
<td>6.35</td>
<td>significant at .05</td>
</tr>
<tr>
<td>F₀ Range</td>
<td>A</td>
<td>5.13</td>
<td>significant at .05</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>3.85</td>
<td>none</td>
</tr>
<tr>
<td></td>
<td>A×B</td>
<td>7.50</td>
<td>significant at .05</td>
</tr>
<tr>
<td>Relative Intensity</td>
<td>A</td>
<td>0.18</td>
<td>none</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>0.11</td>
<td>none</td>
</tr>
<tr>
<td></td>
<td>A×B</td>
<td>0.84</td>
<td>none</td>
</tr>
<tr>
<td>Jitter</td>
<td>A</td>
<td>4.99</td>
<td>significant at .05</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>1.65</td>
<td>none</td>
</tr>
<tr>
<td></td>
<td>A×B</td>
<td>7.92</td>
<td>significant at .05</td>
</tr>
<tr>
<td>Rate of Speech</td>
<td>A</td>
<td>7.51</td>
<td>significant at .05</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>1.21</td>
<td>none</td>
</tr>
<tr>
<td></td>
<td>A×B</td>
<td>1.18</td>
<td>none</td>
</tr>
</tbody>
</table>

*Factor A denotes type of device, B denotes brand.

According to Barney, Haworth, and Dunn (1959), the prototype of the Western Electric 5 A model had a 0.5 millisecond duty cycle to generate sufficient acoustic output with minimal current drain. The device operated at a frequency of 100 pulses per second. Because jitter is frequency modulation, a periodic signal, such as that from an electrolarynx, would exhibit no jitter. In pilot testing, this was not found to be the case. The periodic tone possibly is disturbed by neck or mouth coupling. It was planned, therefore, to collect exploratory pitch perturbation data. As high jitter values correlate with perceived vocal "roughness" in normal laryngeal voices, so may jitter values of artificial larynx speech. Intuitively, this seems supported in the perceptual literature in that listeners rated "quality," as the salient feature differentiating three brands of artificial larynx devices (Bennett & Weinberg, 1973).

Rate of speech. The amount of time needed for each subject to read the 98-word "Rainbow Passage" was measured with a stopwatch (Delta Impex, #1082P) and calculated as words per minute.

RESULTS

The data were analyzed statistically to compare types (factor A) and brands (factor B) of artificial devices with respect to the five acoustic measures. A series of five nested two-way repeated-measures analysis of variance (ANOVA) procedures with replication were completed on the data. Composite results are in Table 2; four of the five measured proved significant.

Reliability of measures also was determined. For the investigation, the average fundamental frequency, fundamental frequency range, and relative intensity were measured during the second sentence of the "Rainbow Passage." To determine the representativeness of these extracted values, three speech samples were selected at random. Each of the three acoustic measures were then made for the entire passage,
sentence by sentence, using the Visi-Pitch walking mode. Averages for the passage were then calculated and reliability between score pairs found. Values extracted from the second sentence of the "Rainbow Passage" were highly representative of the acoustic measures of the entire passage. Specifically, the reliability of fundamental frequency measures was .92. For fundamental frequency range, measurement reliability was .89, and for relative intensity it was .90.

Speech rates were measured twice for all 56 samples. When stopwatch measures were in disagreement, an average reading time was taken as the criterion measure. No remeasures were necessary for jitter from the prolonged /i/ as determined by the Visi-Pitch/Apple II E subroutine.

**Fundamental Frequency**

The fundamental frequency exhibited during the reading of the second sentence of the "Rainbow Passage" differed significantly among device brands \( F = 5.87, df = 1.13, p < .05 \). The interaction of type and brand of device was also significant \( F = 6.35, df = 1.13, p < .05 \) yet the main effect of type was not. The data help illuminate these findings. In general, fundamental frequency of artificial devices is similar whether one uses a neck type or intraoral type: \( F_o \) of speakers with the intraoral devices averaged 80.67 Hz \( [SD = 15.55] \) and with the neck placement devices it averaged 79.67 Hz \( [SD = 16.75] \).

Post hoc testing using the Tukey A showed only one of the multiple comparisons to be significant at the .05 level. Only the brand differences are of interest. The mean \( F_o \) of the oral Servox device, at a high of 89.08 Hz \( [SD = 17.36] \), differed significantly from the low 72.26 Hz value \( [SD = 6.37] \) of the Cooper-Rand. Mean fundamental frequency of the other brands were similar and nonsignificant: neck Servox 79.89 Hz \( [SD = 16.67] \), AT&T 79.45 Hz \( [SD = 16.83] \).

**Fundamental Frequency Range**

Fundamental frequency range was a measure of intonation used by speakers within the second sentence of the "Rainbow Passage." Inflectional variability using neck type devices was significantly different than that displayed with intraoral devices \( F = 5.13, df = 1.13, p < .05 \). Specifically, the neck devices allowed an average frequency range of 153.83 Hz, whereas oral devices averaged a 99.13 Hz inflectional range.

The interaction of type and brand proved significant \( F = 7.50, df = 1.13, p < .05 \) although the main effect of brand did not. Tukey A post hoc testing used a .05 level of confidence. The Servox, with a \( F_o \) range averaging 175.61 Hz, was found to exhibit significantly more inflection than the Cooper-Rand, which had the lowest average range of 70.58 Hz. Other devices were comparable in fundamental frequency range. The AT&T model exhibited 132.05 Hz range while the oral-adapted Servox averaged 127.67 Hz.

**Relative Intensity**

The intensity of speech did not differ significantly among types nor brands of artificial devices. Intensity measures (relative to 30 dB) were comparable with all

_Moffet & Pindzola: Acoustic Properties_ 35
neck devices averaging 59.0 dB and all oral devices 58.61 dB. Relative intensities for each of the four brands were: Cooper-Rand 58.26 dB [SD = 4.97], oral Servox 58.97 [SD = 6.56], AT&T 59.76 [SD = 4.42], and Servox 58.23 [SD = 4.76].

**Jitter/Pitch Perturbation**

The cycle-to-cycle variation in vibratory frequency during the prolongation of the vowel /i/ proved significantly different between intraoral and neck type devices \( F = 4.99, df = 1.13, p < .05 \). Oral types of artificial larynges averaged 10.39% [SD = 7.82] of jitter while neck types exhibited more jitter at 17.27% [SD = 8.75].

The interaction of type and brand was significant \( F = 7.93, df = 1.13, p < .05 \) but the main effect of brand was not. In Tukey A testing, multiple comparisons proved significant at the .05 level; only those concerning specific brands are of interest. The Cooper-Rand has significantly less jitter than the Servox and AT&T devices but is comparable with the oral Servox. The oral Servox has significantly less jitter than the AT&T but is similar to the neck Servox. An ordering of the means by brand further illustrates differences in amount of pitch perturbation. The Cooper-Rand exhibits the least jitter at 8.80% [SD = 8.06], followed by the oral Servox at 11.98% [SD = 7.24], neck Servox at 15.74% [SD = 6.77], and AT&T the highest at 18.81% [SD = 10.12].

**Rate of Speech**

Rate of speech was significantly affected by type of device used but not by brand. Likewise, there was no significant interaction. Speech with neck type devices was significantly faster than speech with intraoral devices \( F = 7.51, df = 1.13, p < .05 \) although the mean values appear close. Intraoral device rates averaged 126.57 words per minute (wpm) [SD = 24.39] and neck device rates averaged 124.29 wpm [SD = 27.31]. These large standard deviation values reflect much dispersion among the reading rates with values ranging from 51 to 165 wpm.

Although not significantly different, speech rates per device were as follows: oral Servox 125.14 wpm [SD = 23.61], Cooper-Rand 128.00 wpm [SD = 25.04], neck Servox 132.57 wpm [SD = 25.83], AT&T 136.00 wpm [SD = 28.60].

**DISCUSSION**

Fundamental frequency means of neck and intraoral devices were similar and nonsignificant, according to analysis of variance procedures, yet this is deceiving. Significant differences were found between the Cooper-Rand and oral Servox devices for mean fundamental frequency after post-hoc testing for the interaction effect. In essence, their average fundamental frequencies balanced each other out, making fundamental frequency means of neck and oral devices appear similar. The oral Servox had the highest fundamental frequency mean of 89 Hz, while the Cooper-Rand had the lowest fundamental frequency mean of 72 Hz. Both neck type devices displayed fundamental frequency means of approximately 79–80 Hz. The present results are similar to, yet slightly lower than the 100 Hz average reported in the literature (Barney, Haworth, & Dunn, 1959).
Because research (Rothmann, 1978) shows higher frequency alaryngeal speech to be rated more superior than lower frequency speech, the present findings have clinical application. The Servox, whether orally adapted or used with neck placement, has the higher fundamental frequency and, so, would be the device of choice. In this light, Cooper-Rand recommendations would appear tenuous, especially for female laryngectomies.

Mean fundamental frequency comparisons of artificial larynx speech with other speech forms are also of clinical interest. Esophageal speakers had a mean fundamental frequency of 77.1 Hz in the Robbins et al. (1984) study. Houseman (1987) found esophageal speakers to average 84 Hz. Therefore, fundamental frequencies of electronic artificial larynx devices are similar to or slightly higher than fundamental frequencies of esophageal speech. Also, in both of these studies, mean fundamental frequencies for normal \( x = 103 \) Hz and 92 Hz and tracheoesophageal \( x = 77 \) Hz and 108 Hz speakers were higher than the fundamental frequencies of the artificial devices in the present study.

The data on fundamental frequency range suggest that intonation variability is possible with both neck and intraoral devices. The statistical superiority of neck devices is, however, somewhat surprising in that the literature often suggests that more inflection is possible with intraoral devices (c.f. Weinberg & Riekena, 1973). Caution must be exercised in such a broad literature interpretation; what is meant is that pneumatic intraoral devices, especially the Tokyo, are capable of producing greater intonational variability. With the electronic devices it is wholly plausible, from a design point of view, that neck coupling allows for more fundamental frequency variability than does intraoral tube placement. Indeed, clinicians often train patients to vary the amount of pressure used in neck contact for the regulation of pitch, loudness, and mechanical noise.

The Servox had the greatest range of fundamental frequency, averaging 176 Hz during an intoned sentence. The Cooper-Rand, with only a 71 Hz range, exhibited the least amount of inflection of the devices tested. These data are consistent with the 100–200 Hz range reported by Weiss, Yeni-Komshian, and Heinz (1979) for the Western Electric 5 A electrolarynx. This information should be remembered when counseling patients and selecting devices.

The range of the Servox is comparable to the 177 Hz range of esophageal speakers reported by Houseman (1987). The fundamental frequency range of 150 Hz for tracheoesophageal speakers, in her study, is lower than the Servox’s range in the present investigation.

The intensity of speech produced by all four devices was similar with values in the 58–59 dB range (± 30 dB internal referent). Clinically, this can be interpreted as meaning that the four neck and intraoral electrolarynges are neither superior nor inferior to each other. All may be recommended to patients with confidence, regarding speech output loudness.

The results of the pitch perturbation data were surprising in that electrolarynges had high amounts of jitter despite a supposed periodic duty cycle. Perhaps the transmission or conduction of the sound through tubes or tissues imparts an irregularity to the pulses. The fact that intraoral devices had less of this perturbation than neck devices may suggest that oral tubes distort the periodic duty cycle less.
An alternative hypothesis is that the oral cavity absorbs or muffles distortions in the wave more than neck tissues and so the output spectrum contains less perturbation.

It was not surprising that the neck devices were associated with significantly faster rates of speech than the intraoral devices. Clinical opinion that an oral tube slows articulation rates therefore seems supported. Intraoral device rates, averaging 127 wpm, and neck device rates, averaging 134 wpm, are somewhat slower than the 187 wpm rate reported in the literature for electrolarynges (Goldstein & Rothman, 1976). The present data, however, are similar to the 138 wpm rate reported by Weinberg and Riekena (1973) for the Tokyo pneumatic device.

Comparatively, artificial larynx speech in the 125–136 wpm range is faster than esophageal speech rates as reported by Robbins et al. (1984) and Houseman (1987). They reported means of 99 and 94 words per minute, respectively. The present artificial larynx device means are similar to the 128 wpm reported for tracheoesophageal speakers by Robbins et al. (1984), but are much lower than the 152 wpm rate reported by Houseman (1987). Both studies found normal speech to range 170–173 words per minute which is much higher than the rate of artificial larynx speech.

CONCLUSIONS

The preceding discussion elaborated on acoustic differences and similarities among the four electrolarynges tested. What is not clear is the extent of impact the individual user has on the devices. For example, although an artificial device may be preset at a given frequency, the fundamental frequencies measured on 14 subjects were not identical. Individual factors, such as those associated with tissue coupling, affected the data obtained “from the instrument.” Even so, the findings in this study do reflect basic differences (or differences in use) among the four devices because such trends emerged from a group of subjects unfamiliar with particular devices.

It is of clinical importance to know the relative merits of the devices to better assist patients in their selection process. To this end, a somewhat subjective interpretation of the mean acoustic findings has been attempted (Moffet, 1987; Rothman, 1978). Table 3 organizes the acoustic information concerning mean differences; the brands are ordered from “best to worst” to serve as a quick reference for clinicians.

REFERENCES


38 NSSLHA Journal


