

## Research Article

# Variability of Electrolaryngeal Speech Intelligibility in Multitalker Babble

2003

 Steven R. Cox,<sup>a</sup>  Kimberly McNicholl,<sup>a</sup> Christine H. Shadle,<sup>b</sup> and Wei-rong Chen<sup>b</sup>

**Purpose:** The purpose of this study was to report the variability of electrolarynx (EL) users' speech intelligibility in quiet and in multitalker babble.

**Method:** Ten EL users (five Servox® Digital, five TruTone™) who were at least 2 years postlaryngectomy provided recordings of five sentences from the 1965 Revised List of Phonetically Balanced Sentences. Recordings were judged by two groups of naïve listeners in quiet and in the presence of multitalker babble. Fifteen listeners orthographically transcribed a total of 750 sentences containing 3,750 key words in quiet, and another 15 listeners orthographically transcribed the same sentences mixed with multitalker babble.

**Results:** Significant differences in speech intelligibility were observed between listening conditions; 17.9% more key words were correctly identified in quiet compared to

multitalker babble. Significant differences in fundamental frequency ( $F_0$ ) standard deviation and range but not speech intelligibility were observed between EL device types. A positive correlation of moderate significance was observed between  $F_0$  standard deviation and intelligibility for TruTone users in multitalker babble.

**Conclusions:** Findings suggest that listeners are able to identify a significantly higher percentage of EL users' speech in quiet compared to multitalker babble, but a large variability in EL users' speech intelligibility exists. Continued investigation involving a larger number of EL users is necessary to confirm this study's findings. Future research should explore the relationships among  $F_0$  measures, speaker characteristics (e.g., rate of speech, articulatory precision), and speech intelligibility, in addition to improving alaryngeal rehabilitation training protocols for EL users.

The American Cancer Society (2020) estimates 12,370 new laryngeal cancer diagnoses in the United States in 2020. Recent estimates suggest that approximately 4,000 total laryngectomies are performed each year in the United States (Gourin et al., 2019). While the use of total laryngectomy has been declining in the United States (e.g., a decrease of 27.3 cases per year; Orosco et al., 2013), this procedure remains an important primary or salvage treatment for laryngeal disease (Silverman et al., 2019). The electrolarynx (EL) remains a highly usable primary or backup communication method for individuals postlaryngectomy. It is relatively easy to learn and use with appropriate alaryngeal voice and speech rehabilitation (Doyle, 2005; Nagle, 2019). Estimates suggest that EL device use can vary from 30% to 85% at 1 year postlaryngectomy;

however, more recent data suggest that 50% of laryngectomees use an EL up to 5 years postlaryngectomy (Bhandare et al., 2013; Hillman et al., 1998; Mendenhall et al., 2002; Ward et al., 2003).

EL speech is often characterized by a monotonous and "unnatural" vocal quality involving numerous acoustic deficits (i.e., reduced frequency variation, lack of low spectral energy, and radiating device noise; Doyle & Eadie, 2005; Meltzner & Hillman, 2005; Watson & Schlauch, 2009). These deficits often result in poor listener reactions, in addition to decreases in speech intelligibility and other auditory-perceptual judgments (Bennett & Weinberg, 1973; Cox & Doyle, 2018; Evitts & Searl, 2006; Weiss & Basili, 1985). For example, EL users have been reported to have a wide variability in speech intelligibility with an approximate range of 16%–90% (Cox, 2019). This variability can be attributed to an EL device's fundamental frequency ( $F_0$ ) setting, lack of frequency variation, lack of distinction between voiced and unvoiced phonemes, and reduced low-frequency energy in the source spectrum (Gandour & Weinberg, 1984; Goldstein & Rothman, 1976; Laures & Weismer, 1999; Nagle et al., 2012; Watson & Schlauch, 2009). Nagle et al. (2012) found that EL devices with a lower  $F_0$  (e.g., 75 Hz) facilitated higher speech intelligibility

<sup>a</sup>Department of Communication Sciences and Disorders, Adelphi University, Garden City, NY

<sup>b</sup>Haskins Laboratories, New Haven, CT

Correspondence to Steven R. Cox: scox@adelphi.edu

Both Steven R. Cox and Kimberly McNicholl are primary authors.

Editor-in-Chief: Julie Barkmeier-Kraemer

Editor: Rita R. Patel

Received April 15, 2020

Revision received June 8, 2020

Accepted June 29, 2020

[https://doi.org/10.1044/2020\\_AJSLP-20-00092](https://doi.org/10.1044/2020_AJSLP-20-00092)

**Disclosure:** The authors have declared that no competing interests existed at the time of publication.

scores compared to devices with a higher  $F_0$  (e.g., 130 or 175 Hz). Concerning frequency variation, Laures et al. (Laures & Bunton, 2003; Laures & Weismer, 1999) compared speech with varying intonation to that with a flattened mean  $F_0$  and found that varying intonation resulted in higher speech intelligibility in background noise. These findings were further demonstrated in EL speech by Watson and Schlauch (2009), who investigated the relationship between frequency variation and speech intelligibility. EL users' speech intelligibility was at least 10% higher when speakers used an EL device with variable frequency control (e.g., a TruTone EL) compared to a flattened frequency (e.g., a Servox Digital EL). This is an important finding since a noisy communication environment presents considerable challenges to both EL users and their communication partners. This is especially true when considering that an EL user's speaking proficiency is often measured using speech intelligibility. Unfortunately, the effect of noise on EL users' speech intelligibility has received limited attention (Clark, 1985; Clark & Stemple, 1982; Holley et al., 1983).

Clark and Stemple (1982) investigated the speech intelligibility of sentences produced by EL users (Servox) and laryngeal, esophageal (ES), and tracheoesophageal (TE) speakers in a variety of listening conditions. Twenty listeners transcribed sentences presented with signal-to-noise ratios (SNRs) of 0, -5, and -10 dB. While no differences in speech intelligibility were found among the four speech modes at 0 dB, listeners identified more sentences produced by EL users in the -5- and -10-dB noise conditions. This suggests that listeners might understand EL users more than ES or TE speakers when background noise is present. However, it is important to note that the researchers used a different speaker for each method of alaryngeal communication, and therefore, there were no data provided regarding the variability in individual EL user performance in noise.

Holley et al. (1983) examined one alaryngeal speaker who used ES and EL speech, in addition to a laryngeal speaker for comparison. Listeners were presented with sentences spoken in quiet and with multitalker babble noise added to create SNRs of +3 and -1 dB. Results revealed that there were significant reductions in intelligibility of all three speech modes as noise increased. When using an EL, the alaryngeal speaker had approximately 85% sentence intelligibility in quiet, approximately 70% sentence intelligibility for the +3-dB SNR condition, and approximately 40% for the -1-dB SNR condition (Holley et al., 1983). When using ES, the alaryngeal speaker had approximately 94% sentence intelligibility in quiet, approximately 70% sentence intelligibility for the +3-dB SNR condition, and approximately 50% for the -1-dB SNR condition (Holley et al., 1983). Therefore, the alaryngeal speaker was less intelligible when using an EL compared to ES in quiet and -1-dB SNR conditions but had similar intelligibility scores in the +3-dB SNR condition. Furthermore, the alaryngeal speaker had lower intelligibility when using an EL than the laryngeal speaker in the quiet condition only; the laryngeal speaker had approximately 100% sentence intelligibility in quiet, approximately 70% sentence intelligibility for the

+3-dB condition, and 30% sentence intelligibility for the -1-dB condition. The most intriguing finding was that there was no significant difference in sentence intelligibility between the three modes of speech in the +3-dB SNR condition. The researchers also found that there was a statistically significant improvement observed in EL speech and ES in the -1-dB SNR condition compared to laryngeal speech. Holley et al. explained that these findings were potentially the result of the acoustic similarities between the laryngeal speech and multitalker babble, whereas there would have been greater perceptual contrasts with alaryngeal speech modes. EL speech, then, may provide some unique benefits in competing noise conditions due to the acoustic and perceptual differences between EL and laryngeal speech. However, the researchers acknowledged that further research must be conducted to provide data regarding the variability of individual EL user performance.

Clark (1985) investigated younger and older listeners' perception of laryngeal speech, EL speech, ES, and TE speech in noise. Sentences were presented at SNRs of 0, -5, and -10 dB to two groups of listeners: 11 younger listeners with a mean age of 27 years (range: 21–30 years) and 11 older listeners with a mean age of 57 years (range: 50–66 years). Younger listeners identified 100% of sentences spoken by the EL user in the 0-dB SNR condition, 91.81% in the -5-dB SNR condition, and 38.18% in the -10-dB SNR condition. Older listeners identified 100% of sentences spoken by the EL user in the 0-dB SNR condition, 85.45% in the -5-dB SNR condition, and 25.45% in the -10-dB SNR condition. Findings suggested that EL speech was the most intelligible alaryngeal method of communication in all three conditions. When comparing laryngeal and EL user performance, no significant differences were noted in 0 and -5 dB SNRs. However, EL users' intelligibility was higher in the -10-dB SNR condition than laryngeal talkers. Clark performed an acoustic analysis of the noise signal and each speaker's sentences and found that the energy in the frequency spectrum of laryngeal, ES, and TE speech was concentrated below 600 Hz, whereas the EL signal remained strong up to 1400 Hz. Clark stated that, "the auditory competition provided less background masking interference for the artificial larynx speech signal than for the other speech signals" (p. 65).

More recent efforts examining the impact of noise on the speech intelligibility of alaryngeal speech have been conducted by Eadie et al. (2016). Their study examined the effect of noise on intelligibility of TE speakers and self-reported quality of life outcomes. Twenty-four TE speakers ( $M_{\text{age}} = 64$  years, range: 39–86 years) at least 1 year post-laryngectomy were recorded while reading sentences, in addition to completing self-reported quality of life questionnaires. Sentences were transcribed by 66 inexperienced listeners ( $M_{\text{age}} = 24$  years, range: 19–45 years): One group of 33 listeners transcribed sentences in quiet, and another group of 33 listeners transcribed sentences in noise (i.e., multitalker babble) with a +6 dB SNR. Findings suggested that TE speakers were more intelligible in quiet (average intelligibility was 93.27%) than in noise (average

intelligibility was 68.64%). Furthermore, noise significantly impacted self-reported measures of quality of life. Eadie et al. (2016) suggested that speech intelligibility in noise might best serve to index self-reported communicative function, especially for those who demonstrate higher speech intelligibility scores in noise.

Holley et al. (1983) stated that “[i]t should be noted that there is considerable variability in the speaking abilities of laryngectomized individuals” (p. 155). This is true for EL users in quiet, but the few studies that have examined EL users’ speech intelligibility in noise were based on judgments of a single EL user. Understanding the variability of multiple EL users’ speech intelligibility in multitalker babble will provide speech-language pathologists (SLPs) with potential alaryngeal voice and speech rehabilitation targets based on EL user performance in everyday communication contexts. The purpose of this study, then, was to report the variability of EL users’ speech intelligibility in quiet and in multitalker babble. The following research questions were addressed:

1. Is the speech intelligibility of 10 EL users significantly different in multitalker babble compared to quiet?
2. Is there a significant difference between Servox Digital and TruTone users’ speech intelligibility in quiet and multitalker babble?
3. Are  $F_0$  characteristics correlated with intelligibility?

## Method

### *Speech Stimuli Recording*

#### Speakers

Speech samples from 10 male EL users were obtained from an archival database (Cox, 2016). The same EL users served as speakers in Cox et al. (Cox & Doyle, 2018; Cox et al., 2019); while the previous studies investigated clear speech in EL users, the current study addressed speech intelligibility in multitalker babble using conversational (or “habitual”) speech. All EL users were recruited at an International Association of Laryngectomees meeting and responded to study advertisements. EL users ranged in age from 59 to 87 years ( $M = 74$ ), and their primary language was English. EL users reported to be in good general health at the time of the study, with no known neurological, medical, or psychological conditions. A neck-type EL was their primary method of communication, and they were at least 2 years postlaryngectomy ( $M = 11$  years,  $SD = 7.3$ , range: 2–19 years) at the time of recording. Each laryngectomee brought their own EL device for the experimental recording session; this included an equal representation of five individuals who used a Servox Digital EL and five individuals who used a TruTone EL. Cox and Doyle (2018) confirmed that all EL users were proficient as a result of using an EL device for at least 2 years postlaryngectomy, and they passed a preliminary intelligibility assessment. Informed consent was obtained from all EL users at the

beginning of the recording session (Western University Research Ethics Board Approval 105382).

#### Speech Stimuli

Ten lists containing sentences from the 1965 Revised List of Phonetically Balanced Sentences were prepared for each speaker (Rothauer et al., 1969). Each sentence contained five key words and ranged from seven to 12 words in length. Four of the key words in each sentence were monosyllabic, and one key word was bisyllabic. These sentences were used due to their low level of predictability. In total, 250 key words (10 speakers  $\times$  5 sentences per speaker  $\times$  5 key words per sentence) were used in the following study for intelligibility scoring purposes.

Speech stimuli recordings were gathered in a quiet room free of background noise as perceptually judged by the first author (S. R. C.). A Shure PG-81 microphone was attached to a desktop microphone stand, and a mouth-to-microphone distance of 15 cm from each speaker’s mouth was maintained. All speaker stimuli were recorded onto a laptop computer using a sampling rate of 44.1 kHz using the SonaSpeech II software employing the Multidimensional Voice Profile application (Kay Pentax). All speech samples were saved on the computer and converted into .WAV files (Audacity 2.2.2; Mazzoni & Dannenberg, 2018). The recordings of sentences were obtained as part of a larger protocol that included the rainbow passage, a list of 18 monosyllabic words, and 10 sentences from the 1965 Revised List of Phonetically Balanced Sentences per each EL user. Each EL user repeated this procedure in clear speech after reading sentences in habitual speech (Cox, 2016; Cox & Doyle, 2018; Cox et al., 2019).

### *Intelligibility Assessment*

#### Listeners

Thirty women enrolled in undergraduate or graduate studies participated in the intelligibility assessment. Listeners had a mean age of 20 years (range: 18–23 years). All listeners passed a pure-tone hearing screening for the frequencies 500, 1000, 2000, and 4000 Hz at 25 dB HL in each ear. All listeners were monolingual, and their primary language was American Standard English. They reported no history of speech, language, and/or hearing deficits. Listeners were considered to be “naïve” because they verbally confirmed that they had not received training in voice disorders and they had not previously participated in research involving voice disorders. Listeners were not reimbursed for their participation. All listener procedures and recruitment were approved by the primary researchers’ institutional review board (IRB 081117).

#### Listener Stimuli

The first five sentences from each speaker’s recordings were selected and normalized to 65 dB SPL using Praat (Version 6.0.38; Boersma & Weenink, 2018). Five hundred milliseconds of silence were added to the beginning and end of each file using Audacity (Version 2.2.2;

Audacity Team, 2020). These 50 edited sentences were copied into two folders: “Quiet” and “Babble.” Each .WAV file in the Babble folder was mixed with multitalker babble produced by one male speaker and three female speakers (QuickSIN Speech-in-Noise Test; Etymotic Research, 2006). The multitalker babble (henceforth known as “babble”) was normalized to 59 dB SPL using a normalization script in Praat to ensure an SNR of +6 dB was achieved between sentences and babble. Each sentence in the babble condition was edited to ensure that the first 500 ms contained babble only, followed by babble and the recorded sentence, and ended with 500 ms of babble (Eadie et al., 2016; Van Engen & Bradlow, 2007).

The Quiet and Babble folders were reduplicated until 15 folders, each containing 50 sentences, were created for each listening condition. The sentences in the Quiet and Babble folders were then separately randomized. Ten sentences (20%) from each list of 50 sentences were randomly selected as reliability samples. The same 10 reliability samples in quiet were added to the end of the randomized 50 sentences in each Quiet folder, and the same 10 reliability samples in babble were added to the end of the randomized 50 sentences in each Babble folder. Five familiarization samples of an EL user who was not evaluated for intelligibility were added to the beginning of all folders. The EL user read five unique sentences from a list in the 1965 Revised List of Phonetically Balanced Sentences. These samples were presented to all listeners without babble and were used to acquaint listeners with EL speech prior to the listening procedure.

### Listening Procedure

Listeners were randomized into two groups that were matched by age: One group of 15 women ( $M_{\text{age}} = 21$  years, range: 18–24 years) listened to sentences spoken in quiet, and another group of 15 women ( $M_{\text{age}} = 21$ , range: 18–23 years) listened to sentences spoken in babble. Presentation of each listening condition was counterbalanced across listeners; for example, Listener 1 made intelligibility judgments in quiet, and Listener 2 made intelligibility judgments in babble. After informed consent was obtained, all listeners were provided with instructions similar to those described by Eadie et al. (2016):

You will be listening to adult speakers who have had total removal of their voice box due to cancer. These speakers are using a method of speech called “electrolaryngeal speech.” We are interested in how well listeners can understand these speakers in both quiet and background noise. You will only hear samples presented in quiet or noise. We will play some sentences, and we would like you to write out the words that you hear. You may listen to the sentences up to 2 times. Some of these sentences will be difficult to understand. Do your best, and guess when you need to. You may listen to each sentence 2 times (p. 397).

Since Eadie et al. (2016) examined intelligibility in noise (i.e., babble) using TE speakers, modifications were

made to inform listeners that they were going to listen to “electrolaryngeal speech” in this study.

Listeners were provided with transcription sheets numbered 1–60, with a blank space beside each number. They began each session by listening to five familiarization samples, and then each listener proceeded to click the .WAV file of each stimulus sample and orthographically transcribed what they heard. All stimuli were presented through headphones (Shure SRH440) in a quiet environment as perceptually judged by the first author (S. R. C.). Overall, listening sessions required an average of 26.9 min ( $SD = 4.8$ ) in quiet and 27.5 min ( $SD = 4.9$ ) in babble.

### Statistical Analysis

#### Reliability Analyses

Intrarater reliability for speech intelligibility was assessed using Pearson product-moment correlation coefficients for each of 30 listeners using the intelligibility ratings of the repeated 20% of stimuli. Interrater reliability for speech intelligibility was assessed using intraclass correlation coefficients (ICCs) and their 95% confidence intervals based on a two-way mixed-effects model with absolute agreement and a mean of  $k$  raters, (i.e., ICC(2,  $k$ ) model; Shrout & Fleiss, 1979).

#### Acoustic Analyses

Every sentence was pitch-tracked using autocorrelation analysis on Praat with a 50-ms window and a manually determined search range. Characteristics of  $F_0$  were then assessed for each EL user:  $F_0$  mean, standard deviation, minimum and maximum, and range. The robust estimations of minimum and maximum  $F_0$ s were defined as the 0.5% and 99.5% quantiles, respectively, of the  $F_0$  values in all sound files for each EL user. The user-specific  $F_0$  ranges were then calculated as the differences between the estimated minimum and maximum  $F_0$ s. Both  $F_0$  standard deviation and  $F_0$  range were measures of  $F_0$  variability and calculated separately for each sentence and then averaged across sentences for each user. The very small but nonzero  $F_0$  ranges for the Servox Digital users in the results confirmed the robustness of our estimates. Multiple independent-samples  $t$  tests were computed to assess a familywise null hypothesis: Servox Digital and TruTone users do not differ in the mean and variability (standard deviation and range) of  $F_0$ . The false discovery rate (Benjamini & Hochberg, 1995) was used at the level of 0.05 to correct the  $p$  values of multiple  $t$  tests.

#### Intelligibility Analyses

Speech intelligibility scores were calculated for each EL user by dividing the number of correctly identified key words by the total number of key words. Key words were considered to be correct if they phonemically matched the target key word, and misspellings were counted as correct (Eadie et al., 2016; Hustad & Cahill, 2003). Proportional change was defined as the difference between intelligibility in babble and quiet divided by intelligibility in quiet. A



paired-samples *t* test was used to assess the effect of listening condition on speech intelligibility. This was followed by multiple comparisons of intelligibility scores within each device group according to listening condition (e.g., Servox Digital users' intelligibility scores in quiet vs. babble, TruTone users' intelligibility scores in quiet vs. babble) and then between device groups and speaking conditions (e.g., Servox Digital users' intelligibility scores in quiet vs. TruTone users' intelligibility score in quiet, Servox Digital users' intelligibility score in babble vs. TruTone users' intelligibility score in babble). An a priori significance level was set at  $p < .05$  for all statistical analyses. The false discovery rate was used to correct the  $p$  values for two familywise null hypotheses: (H1) Babble noise does not have an effect on intelligibility on Servox Digital and TruTone users and (H2) Servox Digital and TruTone users do not differ in intelligibility in both quiet and babble conditions.

### Correlation Analyses

Pearson product-moment correlations were used to assess the relationship between intelligibility and the  $F_0$  characteristics that were significantly different between Servox Digital and TruTone users.

## Results

### Reliability Analysis

Intelligibility judgments of the 20% of the samples that were repeated ( $n = 10$ ) were used for calculating intrarater reliability in quiet and babble conditions. Given that listeners in this experiment had no previous experience in assessing EL speech, in addition to the atypical quality of EL users' speech, a specific criterion of Pearson  $r \geq .5$  was used. In total, two listeners were eliminated from further analyses using this criterion. The mean intrarater reliability for intelligibility judgments pre- and postlistener exclusion is presented in Table 1. Interrater reliability for the 14 listeners providing judgments in quiet was calculated as ICC = .949 (95% CI [0.925, 0.967]) and ICC = .947 (95% CI [0.922, 0.967]) for the 14 listeners providing judgments in babble.

### EL Device Characteristics

Characteristics of EL users' device type and their  $F_0$  characteristics (mean, standard deviation, minimum, maximum, and range) are shown in Table 2. The mean  $F_0$  for EL users with a Servox Digital was 74.8 Hz ( $SD = 16.1$ , range: 46.7–84.7), and the mean  $F_0$  for EL users with a

**Table 1.** Mean intra-rater reliability for judgments of intelligibility without (the center column,  $N = 15$ ) and with (right column,  $N = 14$ ) exclusion on the basis of criterion.

Judgment	Pearson $r$ ( $SD$ )	Pearson $r$ ( $SD$ ) with exclusion
Speech intelligibility in quiet	.85 (.16)	.88 (.11)
Speech intelligibility in babble	.79 (.18)	.82 (.15)

**Table 2.** Electrolarynx user device type and frequency characteristics.

Speaker	Device	Mean $F_0$	$F_0$ $SD$	Min $F_0$	Max $F_0$	$F_0$ range
1	Servox Digital	83.6	0.3	82.8	84.4	1.6
2	Servox Digital	46.7	0.1	46.5	47	0.5
3	Servox Digital	83.4	0.5	80.3	84.5	4.2
4	Servox Digital	75.6	0.2	74.2	76.8	2.6
5	Servox Digital	84.7	0.6	79.5	85.5	6
6	TruTone	77.6	1.1	75	79.8	4.8
7	TruTone	69.2	0.6	67.6	70.6	3
8	TruTone	91.3	2.4	82.3	93.8	11.5
9	TruTone	75.5	3.1	68.7	83.9	15.2
10	TruTone	93.9	3.8	81.5	107	25.5

Note. All data are provided in Hz.  $F_0$  = fundamental frequency.

TruTone was 81.5 Hz ( $SD$  Servox = 10.6, range: 69.2–93.9). Furthermore, Servox Digital users had a mean  $F_0$  range of 2.98 Hz, and TruTone users had a mean  $F_0$  range of 12 Hz. Multiple *t* tests revealed that the Servox Digital users and TruTone users did not significantly differ in  $F_0$  mean (adjusted  $p = .476$ ) but did significantly differ in both  $F_0$  range (adjusted  $p = .0495$ ) and  $F_0$  standard deviation (adjusted  $p = .0495$ ).

### Speech Intelligibility

Speech intelligibility was based on a total of 3,500 perceptual ratings (5 key words  $\times$  5 sentences  $\times$  10 speakers  $\times$  14 listeners) in each listening condition. Speech intelligibility scores were grouped according to "quiet" and "babble" and are shown in Table 3 and Figure 1. EL users had a mean intelligibility of 75.5% ( $SD = 20.8\%$ ; range: 33.1%–92.9%) in quiet and 57.3% ( $SD = 23.3\%$ , range: 12.9%–78.9%) in babble. There was a mean difference of 17.9% between conditions. A paired-samples *t* test revealed a statistically significant difference between speech intelligibility scores in quiet and babble conditions,  $t(9) = 2.262$ ,  $p = 3 \times 10^{-5}$ .

### Comparison of Intelligibility by Condition and Device Group

Table 3 also provides a comparison of speech intelligibility scores between each condition and the EL device used. Servox Digital users had mean intelligibility scores of 73.4% ( $SD = 24.4\%$ , range: 33.1%–92.9%) in quiet and 51.2% ( $SD = 26.2\%$ , range: 12.9%–67.1%) in babble. TruTone users had mean intelligibility scores of 77.6% ( $SD = 19.2\%$ ; range: 44.3%–90.9%) in quiet and 64.0% ( $SD = 21.0\%$ , range: 27.4%–78.9%) in babble. These data indicate TruTone users had intelligibility scores that were 4.2% and 12.8% greater than Servox Digital users in quiet and babble, respectively. No significant differences in intelligibility were found between Servox Digital and TruTone users in quiet ( $p = .77$ , adjusted  $p = .77$ ) or babble ( $p = .42$ , adjusted  $p = .77$ ). The nonsignificance for the seemingly meaningful difference (12.8%, Cohen's  $d = 0.54$ , medium effect size) in intelligibility between Servox Digital and TruTone users in babble is potentially due to low statistical

**Table 3.** Speech intelligibility scores for key words in quiet and babble by device type.

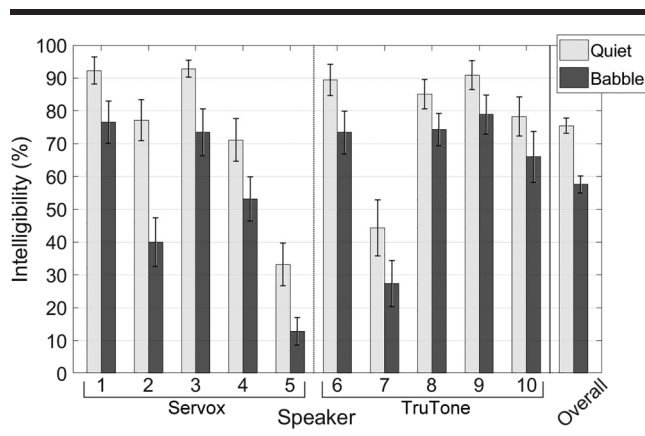
Speaker	Quiet (a)	Babble (b)	Difference = abs (b - a)	Proportional change (%) = (b - a)/a × 100
	%	%	%	
1	92.3	76.6	15.7	-17.0
2	77.1	40.0	37.1	-48.1
3	92.9	73.4	19.4	-20.9
4	71.1	53.1	18.0	-25.3
5	33.1	12.9	20.3	-61.2
Servox Digital mean	73.3	51.2	22.1	-30.2
6	89.4	73.4	16.0	-17.9
7	44.3	27.4	16.9	-38.1
8	85.1	74.3	10.9	-12.8
9	90.9	78.9	12.0	-13.2
10	78.3	66.0	12.3	-15.7
TruTone mean	77.6	64.0	13.6	-17.5
Overall mean	75.5	57.6	17.9	-23.7
SD	20.8	23.3		

Note. A score of 100% corresponds to all keywords judged correct for that speaker (350 judgments = 5 keywords × 5 sentences × 14 listeners).

power. A paired-samples *t* test revealed a significant difference between speech intelligibility scores for quiet and babble conditions for Servox Digital users,  $t(4) = 5.78$ , adjusted  $p = .0045$ ; similarly, there was a significant difference between intelligibility scores in quiet and babble conditions for TruTone users,  $t(4) = 11.5$ , adjusted  $p = .007$ .

Figure 2 represents the distributions of intelligibility scores in standardized box plots. The height of each box indicates the interquartile range of the distribution, calculated with 25 data points (5 EL users × 5 sentences), and the upper and lower whiskers reflect the minimum and maximum values of the distribution, respectively (excluding outliers). As shown in Figure 2d, the difference in proportional changes between the two devices showed that the babble condition affected TruTone users' intelligibility to

**Figure 1.** Overall and individual speech intelligibility scores in quiet and babble. Error bars represent  $\pm 1.96 SE$  as estimate of 95% confidence interval for the mean.

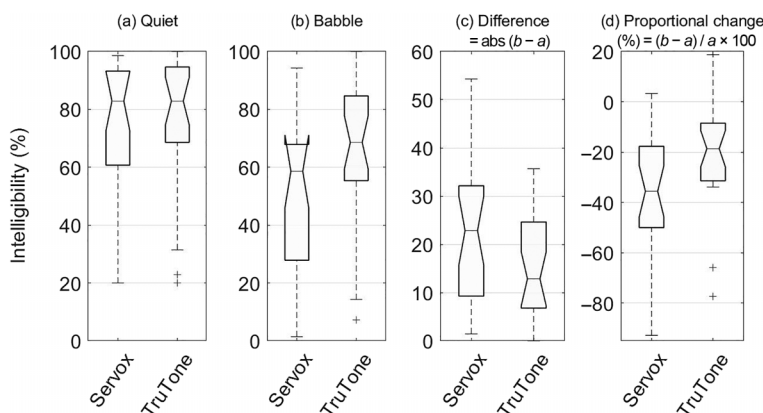


a lesser extent when compared to Servox Digital users. While the difference in proportional change scores between devices was not found to be statistically significant, a difference in proportional change scores of 12.7% might be clinically meaningful. However, the current study did not have the sample size to fully address such analyses.

### Correlation Between $F_0$ Variability and Intelligibility

Two measures of  $F_0$  variability (i.e., range and standard deviation) were found to be significantly different between Servox Digital and TruTone users, and as a result, the correlations between  $F_0$  range and standard deviation with intelligibility were calculated. Figure 3 represents the correlation of  $F_0$  range with intelligibility in quiet (see Figure 3a) and babble (see Figure 3b) and proportional changes from quiet to babble conditions (see Figure 3c). Figure 4 shows the correlations of  $F_0$  standard deviation with intelligibility in the same order. Circles represent Servox Digital users, and squares represent TruTone users. Each circle or square symbol indicates one sentence produced by an EL user. Regression trend lines were calculated only for TruTone users because, theoretically, the variability of  $F_0$  for Servox Digital users should be zero; the small but nonzero values of  $F_0$  range and standard deviation for Servox Digital users were due to measurement noise. There was no significant correlation between  $F_0$  range and intelligibility in any condition. However, there was a moderate, positive correlation between  $F_0$  standard deviation and intelligibility in babble noise condition ( $r = .43$ ,  $p = .03$ ; see Figure 4b) and proportional change ( $r = .45$ ,  $p = .02$ ; see Figure 4c). The explained variances ( $r^2$ ) of these correlations were low (.18 and .20 for babble noise and proportional change, respectively), suggesting poor fits in the

**Figure 2.** Standardized box plots for the distributions of intelligibility scores of electrolarynx (EL) speech in (a) quiet and (b) babble conditions, along with (c) the absolute differences of the scores in the two conditions and (d) the proportional changes. The upper and lower bounds of each box are the third quartile and first quartile, respectively. The height of each box indicates the interquartile range. The upper and lower whiskers indicate the minimum and maximum of the distribution (excluding outliers). The “notch” in the middle of the box represents 95% confidence interval for the median. Note that each box represents the distribution of 25 data points (5 EL users × 5 sentences).



regression models. The statistical inferences based on these correlations should be read with caution.

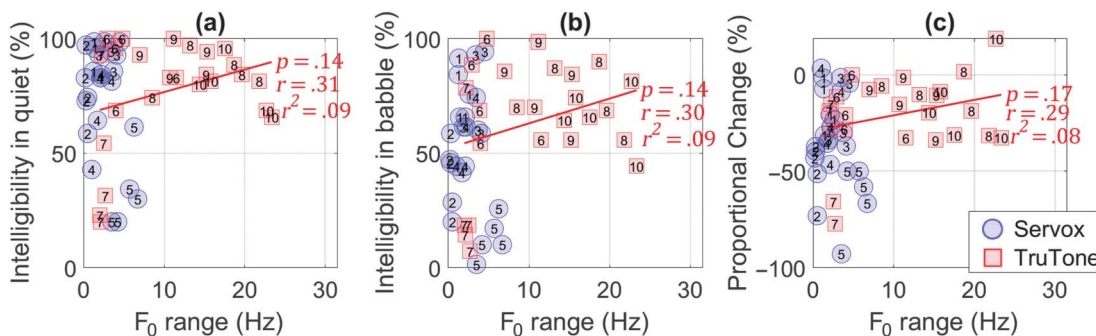
## Discussion

Verbal communication often occurs in environments containing background noise, and this has been shown to reduce speech intelligibility for alaryngeal speakers (Clark, 1985; Clark & Stemple, 1982; Eadie et al., 2016; Holley et al., 1983). Given the dearth of research examining EL users’ speech intelligibility in the presence of background noise, this study reported the variability in speech intelligibility of 10 EL users. Two age-matched groups of 14 normal-hearing listeners made a total of 3,500 judgments in each listening condition. Findings suggest that EL users’ intelligibility scores in quiet and babble conditions were significantly different, and this was also true for each type of EL. However, no significant statistical

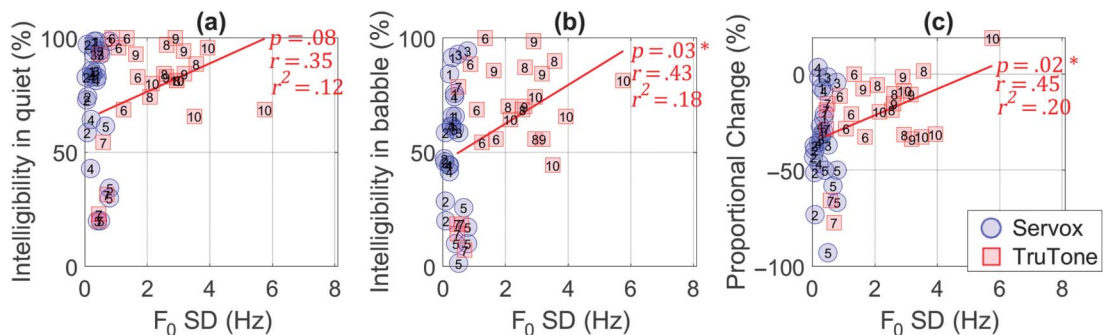
differences were observed when comparing devices within each listening condition.

This study may provide a more representative picture of everyday EL users’ speech intelligibility in quiet and babble when compared to prior research in which only a single EL user was studied (Clark, 1985; Clark & Stemple, 1982; Holley et al., 1983). Holley et al. (1983) identified a difference of 15% between quiet and noise conditions for an EL user; that is, listeners correctly identified approximately 85% of sentences produced in quiet and approximately 70% of sentences produced in noise. The difference between conditions is similar to the present results that suggest a 17.9% difference between quiet and noise (i.e., 75.5% for quiet and 57.6% for noise). Holley et al. used an SNR of +3 dB SNR, which would support the notion of EL users performing similarly to this study involving a +6 dB SNR. However, the current findings are difficult to compare to those by Clark and Stemple (1982) and Clark (1985) for several reasons.

**Figure 3.** Correlation of fundamental frequency ( $F_0$ ) range with intelligibility in (a) quiet, (b) babble, and (c) proportional change from quiet to babble conditions. Circles represent Servox Digital users, and squares represent TruTone users. Numbers within a circle or square indicate user indices; each symbol represents the intelligibility of one sentence spoken by one user. Regression lines represent the correlations of  $F_0$  range and intelligibility for TruTone users only.



**Figure 4.** Correlation of fundamental frequency standard deviation ( $F_0$  SD) with intelligibility in (a) quiet, (b) babble, and (c) proportional change from quiet to babble conditions. Circles represent Servox Digital users, and squares represent TruTone users. Numbers within a circle or square indicate user indices; each symbol represents the intelligibility of one sentence spoken by one user. Regression lines represent the correlations of  $F_0$  SD and intelligibility for TruTone users only.



Clark and Stemple (1982) and Clark (1985) reported speech intelligibility scores of 99.5% and 100%, respectively, for the same EL user in a 0-dB SNR listening condition. In addition, they also reported 95% and 91.81% speech intelligibility, respectively, in a  $-5$ -dB SNR condition. In this study, three out of 10 EL users (two Servox Digital, one TruTone) had speech intelligibility scores of  $>90\%$  in quiet and  $>73\%$  in the  $+6$ -dB SNR condition. The most striking difference, however, is that the three EL users experienced an average decrease of approximately 16% in babble compared to a decrease of approximately 6% in Clark et al. (Clark, 1985; Clark & Stemple, 1982). Two reasons might account for the differences between these studies. First, Clark et al. (Clark, 1985; Clark & Stemple, 1982) reported that EL speech is more intelligible in noisier listening conditions. A listening condition of  $-5$  dB SNR is “noisier” than this study’s  $+6$  dB SNR and, therefore, might have contributed to the higher intelligibility scores in the prior studies. Second, Clark et al. (Clark, 1985; Clark & Stemple, 1982) used the same EL user who was deemed to be “above average” after subjective judgments were made by experienced SLPs. The proficiency of the EL users in the current study was not formally judged by SLPs, but instead, they were required to use an EL as their primary alaryngeal method of communication for at least 2 years postlaryngectomy and passed a preliminary intelligibility assessment (Cox & Doyle, 2018). The findings from this study, then, appear to be more representative of everyday EL users’ speech intelligibility in multitalker babble compared to prior work focused on a “above average” EL user.

Group means according to device groupings showed that TruTone users had higher average speech intelligibility scores compared to Servox Digital users in quiet and babble (i.e., 4.2% and 12.8%, respectively). However, no significant differences were noted when comparing Servox Digital and TruTone users in each condition. This suggests that, similar to the individualized nature of voice and speech rehabilitation, examination of individual data might highlight clinically meaningful changes across listening conditions. For example, even though the TruTone group means

were higher, two Servox Digital users (i.e., Speakers 1 and 3) had the highest overall speech intelligibility scores (i.e., 92.3% for Speaker 1 and 92.9% for Speaker 3) in quiet, and the second and third highest intelligibility scores (i.e., 76.6% for Speaker 1 and 73.4% for Speaker 3) in babble. The findings resulted in absolute differences of 15.7% and 19.4% for Speakers 1 and 3, respectively. One TruTone user (i.e., Speaker 9) had the third highest speech intelligibility score in quiet (i.e., 90.9%) and the highest intelligibility score in noise (i.e., 78.9%), which represents the smallest difference when comparing these three speakers across listening conditions (i.e., 12.0%). Rather than only looking at absolute differences between individual EL users, it appears that SLPs and researchers can also use proportional changes to understand if changes in speech intelligibility are clinically meaningful.

The proportional change data were computed in order to indicate how much of a change EL users experienced with speech intelligibility in babble relative to their performance in quiet (i.e., as a baseline measure). Negative values for the data presented in Table 3 suggest that EL users’ speech intelligibility decreased in babble relative to quiet. The TruTone users’ mean speech intelligibility scores were 4.2% and 12.8% higher than Servox Digital users in quiet and babble, respectively. TruTone users’ speech intelligibility decreased by an absolute difference of 13.6% in quiet, but their proportional change was  $-17.5\%$  in babble relative to their baseline performance (in quiet). Servox Digital users, however, experienced an absolute difference of 22.1% in babble compared to quiet, and the proportional change was  $-30.2\%$  in babble relative to their baseline performance in quiet. As a group, Servox Digital users had a larger proportional decrease by 12.7 percentage points compared to TruTone Digital users. Closer examination of the individual data suggest that the Servox Digital users with highest intelligibility in quiet (e.g., Speakers 1 and 3) actually experienced larger proportional changes (e.g.,  $-17.0\%$  and  $-20.9\%$ , respectively) compared to the most intelligible TruTone user (e.g.,  $-13.2\%$ ). These differences might be explained, in part, by closer examination of



the relationships between  $F_0$  characteristics and speech intelligibility.

Prior research has shown that an  $F_0$  of less than 100 Hz may facilitate improved speech intelligibility for EL users (Nagle et al., 2012). The acoustic analyses in this study revealed that, even though all EL users had a mean  $F_0$  of less than 100 Hz (i.e., approximately 78 Hz), there was considerable variability in speech intelligibility scores in quiet and babble. Furthermore, mean  $F_0$  was not significantly different between device groupings, and therefore, correlation analyses for mean  $F_0$  were not presented. Research has also shown that  $F_0$  variability (e.g.,  $F_0$  standard deviation and range) is one of several factors that can positively impact EL users' speech intelligibility in quiet and noise (Goldstein & Rothman, 1976; Laures & Bunton, 2003; Laures & Weismer, 1999; Nagle & Heaton, 2016; Watson & Schlauch, 2009). Recall that Goldstein and Rothman (1976) reported that "good" EL users (Western Electric No. 5 EL) typically had a mean  $F_0$  range of 16.10 Hz and "poor" EL users had a mean  $F_0$  range of 11.1 Hz (as cited in Rothman, 1978). Servox Digital users, with a constant  $F_0$ , had lower overall intelligibility compared to TruTone users who had a mean  $F_0$  range of 12 Hz. This is in agreement with prior work by Watson and Schlauch (2009), who indicated that greater  $F_0$  ranges might enable EL users to better approximate the intonation patterns of healthy laryngeal speakers. It should be noted, however, that this average  $F_0$  range (e.g., 12 Hz) is closer to that of the "poor" EL users as described by Goldstein and Rothman (1976) and considerably smaller than the  $F_0$  range produced by healthy, laryngeal speakers. This might explain why, even though there was a significant difference between EL device groupings in  $F_0$  range, there was no significant correlation between  $F_0$  range and speech intelligibility scores. Nagle and Heaton (2016) concluded that the pitch modulation capabilities of the TruTone EL might not be used as a result of a lack of skill or device settings (i.e., frequency). This suggests that continued investigation involving a larger number of EL users across a variety of proficiency levels (e.g., "poor" to "good") appears warranted.

A better indicator of  $F_0$  variability during connected speech might be  $F_0$  standard deviation. Larger  $F_0$  standard deviation values reflect speech that is less monotonous. In this study,  $F_0$  standard deviation was found to be significantly different between device groupings and was the only variable significantly (albeit moderately) correlated with TruTone users' intelligibility in noise and proportional change scores.  $F_0$  standard deviation may, in part, contribute to the clinically meaningful difference in TruTone users' intelligibility scores in noise and proportional change scores. This also adds to the current literature suggesting that a variable intonation pattern may contribute toward greater speech intelligibility in noise. While only frequency characteristics were explored in this study, however, a myriad factors contribute to speech intelligibility (e.g., EL user experience and training,  $F_0$  characteristics, rate of speech, articulatory precision). These findings highlight the complex, multidimensional nature of EL voice and speech and

warrant continued research exploring the relationships among  $F_0$  measures, speaker characteristics, and speech intelligibility.

Several limitations within the current study must be acknowledged. First, the speakers and listeners formed homogeneous biological sex groupings. That is, EL users were exclusively male, and listeners were exclusively female. The majority of individuals who undergo laryngectomy are male, so an all-male group of EL users can be considered representative for this client population (e.g., a ratio of approximately four males per one female; American Cancer Society, 2020). EL users also responded to study advertisements and met specific criteria (e.g., spoke American Standard English as their primary language, used an EL as their primary form of alaryngeal communication, and were at least 2 years postlaryngectomy). A majority of the listeners were recruited in a department of communication sciences and disorders, which contains a large percentage of female students (e.g., >90%). In theory, biological sex should not impact speech intelligibility judgments; however, researchers must be mindful of the potential impacts of biological sex on other auditory-perceptual measures (e.g., speech acceptability, listener comfort, speech naturalness). Lastly, while this study was the first to examine speech intelligibility of more than one EL user in noise, incorporating a variety of SNRs and/or frequency characteristics might permit more robust statistical analyses in order to extend the findings from prior research concerning EL users' speech intelligibility in noisy communication environments.

Future research should seek to expand upon this study by exploring the relationship between EL users' speech intelligibility in babble and self-reported quality of life measures (e.g., voice-related quality of life). There is the potential that reduced speech intelligibility in noisy communication environments could lead to reductions in voice-related quality of life for EL users. Research might also include acoustic and perceptual analyses of EL speech in listening conditions with a variety of SNRs and  $F_0$  variability to provide a broader understanding about the auditory-perceptual nature of EL speech in noisy listening conditions. Analyses could focus on the relationship between speech intelligibility and speech acceptability, listener comfort, and/or perceived listener effort. Last but not the least, continued investigation of new training techniques to improve control of variable frequency characteristics of EL devices (i.e., Al-Zanoon et al., 2020; Nagle & Heaton, 2016) appears warranted to enhance this unique population's communication effectiveness in a variety of communication contexts.

## Conclusions

The present findings suggest that listeners are able to identify a significantly higher percentage of EL users' speech in quiet compared to babble, but a large variability in EL users' speech intelligibility exists. TruTone users may derive a clinically meaningful benefit as a result of their ability to vary  $F_0$ . However, the small  $F_0$  variability in the present group of TruTone users, alongside significant but

moderate correlations between  $F_0$  standard deviation and intelligibility, suggest that the results need to be confirmed by studies involving a larger number of EL users. Future research should attempt to understand the potential collinearity among  $F_0$  variability, speaker characteristics (e.g., rate of speech, articulatory precision), and speech intelligibility, in addition to improving alaryngeal rehabilitation training protocols for EL users (i.e., Al-Zanoon et al., 2020; Nagle & Heaton, 2016). Ultimately, this line of research may enhance EL users' speech intelligibility, communication effectiveness, and voice-related quality of life.

## Acknowledgments

This work was supported in part by National Institutes of Health Grant R01DC-002717 to Haskins Laboratories. The authors would like to acknowledge Alexandra Chill, Department of Otolaryngology–Head & Neck Surgery, Mount Sinai–Union Square, New York, NY, for her assistance with participant recruitment.

## References

- Al-Zanoon, N., Parsa, V., & Doyle, P. C. (2020). Using visual feedback to enhance intonation control with a variable pitch electrolarynx. *The Journal of the Acoustical Society of America*, 147(3), 1802–1811. <https://doi.org/10.1121/10.0000936>
- American Cancer Society. (2020). *Cancer facts and figures*. <https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/annual-cancer-facts-and-figures/2020/cancer-facts-and-figures-2020.pdf>
- Audacity Team. (2020). Audacity: Free audio editor and recorder (Version 2.2.2). <https://audacityteam.org/>
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, 57(1), 289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>
- Bennett, S., & Weinberg, B. (1973). Acceptability ratings of normal, esophageal, and artificial larynx speech. *Journal of Speech and Hearing Research*, 16(4), 608–615. <https://doi.org/10.1044/jshr.1604.608>
- Bhandare, N., Morris, C. G., & Mendenhall, W. M. (2013). Voice rehabilitation after total laryngectomy and postoperative radiation therapy. *International Journal of Radiation Oncology, Biology, Physics*, 87(2), S454–S455. <https://doi.org/10.1016/j.ijrobp.2013.06.1200>
- Boersma, P., & Weenink, D. (2018). *Praat 6.0.38: Doing phonetics by computer* [Software]. University of Amsterdam. <https://www.praat.org/>
- Clark, J. G. (1985). Alaryngeal speech intelligibility and the older listener. *Journal of Speech and Hearing Disorders*, 50(1), 60–65. <https://doi.org/10.1044/jshd.5001.60>
- Clark, J. G., & Stemple, J. C. (1982). Assessment of three modes of alaryngeal speech with a synthetic sentence identification (SSI) task in varying message-to-competition ratios. *Journal of Speech, Language, and Hearing Research*, 25(3), 333–338. <https://doi.org/10.1044/jshr.2503.333>
- Cox, S. R. (2016). *The application of clear speech in electrolaryngeal speakers* (Publication No. 3533) [Doctoral dissertation, The University of Western Ontario]. <https://ir.lib.uwo.ca/cgi/viewcontent.cgi?article=5219&context=etd>
- Cox, S. R. (2019). Review of the electrolarynx: The past and present. *Perspectives of the ASHA Special Interest Groups*, 4(1), 118–129. [https://doi.org/10.1044/2018\\_PERS-SIG3-2018-0013](https://doi.org/10.1044/2018_PERS-SIG3-2018-0013)
- Cox, S. R., & Doyle, P. C. (2018). The influence of clear speech on auditory–perceptual judgments of electrolaryngeal speech. *Journal of Communication Disorders*, 75, 25–36. <https://doi.org/10.1016/j.jcomdis.2018.06.003>
- Cox, S. R., Raphael, L. J., & Doyle, P. C. (2019). Production of vowels by electrolaryngeal speakers using clear speech. *Folia Phoniatrica et Logopaedica*, 1–7. <https://doi.org/10.1159/000499928>
- Doyle, P. C. (2005). Clinical procedures for training use of the electronic artificial larynx. In P. C. Doyle & R. L. Keith (Eds.), *Contemporary considerations in the treatment and rehabilitation of head and neck cancer: Voice, speech, and swallowing* (pp. 545–570). Pro-Ed.
- Doyle, P. C., & Eadie, T. L. (2005). The perceptual nature of alaryngeal voice and speech. In P. C. Doyle & R. L. Keith (Eds.), *Contemporary considerations in the treatment and rehabilitation of head and neck cancer: Voice, speech, and swallowing* (pp. 113–140). Pro-Ed.
- Eadie, T. L., Otero, D. S., Bolt, S., Kapsner-Smith, M., & Sullivan, J. R. (2016). The effect of noise on relationships between speech intelligibility and self-reported communication measures in tracheoesophageal speakers. *American Journal of Speech-Language Pathology*, 25(3), 393–407. [https://doi.org/10.1044/2016\\_AJSLP-15-0081](https://doi.org/10.1044/2016_AJSLP-15-0081)
- Etymotic Research (2006). QuickSIN Speech-in-Noise Test (Version 1.3). Etymotic Research, Inc. <https://www.etymotic.com/auditory-research/speech-in-noise-tests/quicksin.html>
- Evitts, P. M., & Searl, J. (2006). Reaction times of normal listeners to laryngeal, alaryngeal, and synthetic speech. *Journal of Speech, Language, and Hearing Research*, 49(6), 1380–1390. [https://doi.org/10.1044/1092-4388\(2006\)099](https://doi.org/10.1044/1092-4388(2006)099)
- Gandour, J., & Weinberg, B. (1984). Production of intonation and contrastive stress in electrolaryngeal speech. *Journal of Speech and Hearing Research*, 27(4), 605–612. <https://doi.org/10.1044/jshr.2704.605>
- Goldstein, L. P., & Rothman, H. B. (1976, November). *Analysis of speech produced with an artificial larynx*. Presented at the American Speech and Hearing Association Convention, Houston, TX, United States.
- Gourin, C. G., Stewart, C. M., Frick, K. D., Fakhry, C., Pitman, K. T., Eisele, D. W., & Austin, J. M. (2019). Association of hospital volume with laryngectomy outcomes in patients with larynx cancer. *JAMA Otolaryngology—Head & Neck Surgery*, 145(1), 62–70. <https://doi.org/10.1001/jamaoto.2018.2986>
- Hillman, R. E., Walsh, M. J., Wolf, G. T., Fisher, S. G., & Hong, W. K. (1998). Functional outcomes following treatment for advanced laryngeal cancer. Part I—Voice preservation in advanced laryngeal cancer. Part II—Laryngectomy rehabilitation: The state of the art in the VA system. Research speech-language pathologists. Department of Veterans Affairs Laryngeal Cancer Study Group. *Annals of Otolaryngology, Rhinology & Laryngology Supplement*, 172, 1–27.
- Holley, S. C., Lerman, J., & Randolph, K. (1983). A comparison of the intelligibility of esophageal, electrolaryngeal, and normal speech in quiet and in noise. *Journal of Communication Disorders*, 16(2), 143–155. [https://doi.org/10.1016/0021-9924\(83\)90045-X](https://doi.org/10.1016/0021-9924(83)90045-X)
- Hustad, K. C., & Cahill, M. A. (2003). Effects of presentation mode and repeated familiarization on intelligibility of dysarthric speech. *American Journal of Speech-Language Pathology*, 12(2), 198–208. [https://doi.org/10.1044/1058-0360\(2003\)066](https://doi.org/10.1044/1058-0360(2003)066)
- Laures, J. S., & Bunton, K. (2003). Perceptual effects of a flattened fundamental frequency at the sentence level under different

- listening conditions. *Journal of Communication Disorders*, 36(6), 449–464. [https://doi.org/10.1016/S0021-9924\(03\)00032-7](https://doi.org/10.1016/S0021-9924(03)00032-7)
- Laures, J. S., & Weismer, G.** (1999). The effect of flattened fundamental frequency on intelligibility at the sentence-level. *Journal of Speech, Language, and Hearing Research*, 42(5), 1148–1156. <https://doi.org/10.1044/jslhr.4205.1148>
- Mazzoni, D., & Dannenberg, R.** (2018). *Audacity 2.2.2* [Software]. Carnegie Mellon University. <http://audacityteam.org/>
- Meltzer, G. S., & Hillman, R. E.** (2005). Impact of aberrant acoustic properties on the perception of sound quality in electrolarynx speech. *Journal of Speech, Language, and Hearing Research*, 48(4), 766–779. [https://doi.org/10.1044/1092-4388\(2005\)053](https://doi.org/10.1044/1092-4388(2005)053)
- Mendenhall, W. M., Morris, C. G., Stringer, S. P., Amdur, R. J., Hinerman, R. W., Villaret, D. B., & Robbins, K. T.** (2002). Voice rehabilitation after total laryngectomy and postoperative radiation therapy. *Journal of Clinical Oncology*, 20(10), 2500–2505. <https://doi.org/10.1200/JCO.2002.07.047>
- Nagle, K. F.** (2019). Elements of clinical training with the electrolarynx. In P. C. Doyle (Ed.), *Clinical care and rehabilitation in head and neck cancer* (pp. 129–143). Springer. [https://doi.org/10.1007/978-3-030-04702-3\\_9](https://doi.org/10.1007/978-3-030-04702-3_9)
- Nagle, K. F., Eadie, T. L., Wright, D. R., & Sumida, Y. A.** (2012). Effect of fundamental frequency on judgments of electrolaryngeal speech. *American Journal of Speech-Language Pathology*, 21(2), 154–166. [https://doi.org/10.1044/1058-0360\(2012\)11-0050](https://doi.org/10.1044/1058-0360(2012)11-0050)
- Nagle, K. F., & Heaton, J. T.** (2016). Perceived naturalness of electrolaryngeal speech produced using sEMG-controlled vs. manual pitch modulation. *Interspeech*, 238–242. <https://doi.org/10.21437/Interspeech.2016-1476>
- Orosco, R. K., Weisman, R. A., Chang, D. C., & Brumund, K. T.** (2013). Total laryngectomy: National and regional case volume trends 1998–2008. *Otolaryngology—Head & Neck Surgery*, 148(2), 243–248. <https://doi.org/10.1177/0194599812466645>
- Rothauer, E. H., Chapman, W. D., Guttman, N., Hecker, M. H. L., Norby, K. S., Silbiger, H. R., Urbanek, G. E., & Weinstock, M.** (1969). IEEE recommended practice for speech quality measurements. *IEEE Transactions on Audio and Electroacoustics*, 17(3), 227–246. <https://doi.org/10.1109/ieeestd.1969.7405210>
- Rothman, H. B.** (1978). Analyzing artificial electronic larynx speech. In S. J. Salmon & L. P. Goldstein (Eds.), *The artificial larynx handbook* (pp. 87–111). Grune.
- Shrout, P. E., & Fleiss, J. L.** (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, 86(2), 420–428. <https://doi.org/10.1037/0033-2909.86.2.420>
- Silverman, D. A., Puram, S. V., Rocco, J. W., Old, M. O., & Kang, S. Y.** (2019). Salvage laryngectomy following organ-preservation therapy—An evidence-based review. *Oral Oncology*, 88, 137–144. <https://doi.org/10.1016/j.oraloncology.2018.11.022>
- Van Engen, K. J., & Bradlow, A. R.** (2007). Sentence recognition in native- and foreign-language multitalker background noise. *The Journal of the Acoustical Society of America*, 121, 519–526. <https://doi.org/10.1121/1.2400666>
- Ward, E. C., Koh, S. K., Frisby, J., & Hodge, R.** (2003). Differential modes of alaryngeal communications and long-term voice outcomes following pharyngolaryngectomy and laryngectomy. *Folia Phoniatrica et Logopaedica*, 55(1), 39–49. <https://doi.org/10.1159/000068056>
- Watson, P. J., & Schlauch, R. S.** (2009). Fundamental frequency variation with an electrolarynx improves speech understanding: A case study. *American Journal of Speech-Language Pathology*, 18(2), 162–167. [https://doi.org/10.1044/1058-0360\(2008\)08-0025](https://doi.org/10.1044/1058-0360(2008)08-0025)
- Weiss, M. S., & Basili, A. M.** (1985). Electrolaryngeal speech produced by laryngectomized subjects: Perceptual characteristics. *Journal of Speech and Hearing Research*, 28(2), 294–300. <https://doi.org/10.1044/jslhr.2802.294>