Neck and Face Surface Electromyography for Prosthetic Voice Control After Total Laryngectomy

Cara E. Stepp, James T. Heaton, Rebecca G. Rolland, and Robert E. Hillman

***Abstract*—The electrolarynx (EL) is a common rehabilitative speech aid for individuals who have undergone total laryngectomy, but they typically lack pitch control and require the exclusive use of one hand. The viability of using neck and face surface electromyography (sEMG) to control the onset, offset, and pitch of an EMG-controlled EL (EMG-EL) was studied. Eight individuals who had undergone total laryngectomy produced serial and running speech using a typical handheld EL and the EMG-EL while attending to real-time visual sEMG biofeedback. Running speech tokens produced with the EMG-EL were examined for naturalness by 10 listeners relative to those produced with a typical EL using a visual analog scale. Serial speech performance was assessed as the percentage of words that were fully voiced and pauses that were successfully produced. Results of the visual analog scale assessment indicated that individuals were able to use the EMG-EL without training to produce running speech perceived as natural as that produced with a typical handheld EL. All participants were able to produce running and serial speech with the EMG-EL controlled by sEMG from multiple recording locations, with the superior ventral neck or submental surface locations providing at least one of the two best control locations.**

***Index Terms*—Biological motor systems, communication aids, electromyography, prosthetics, vocal system.**

I. Background

In 2008, the American Cancer Society estimated that there were 12 250 new cases of laryngeal cancer in the U.S. [1]. In some cases, radical surgical intervention such as total laryngectomy is needed to manage advanced laryngeal cancer. Total laryngectomy consists of removing the larynx, thus removing the natural sound source for speech production. Options for voice rehabilitation after total laryngectomy include esophageal speech, tracheo-esophageal (TE) speech, and the use of an electrolarynx (EL). To produce esophageal speech, an individual must learn to inject air into an esophageal reservoir and to release it through the vibratory pharyngoesophageal segment, a skill that is difficult for many individuals to acquire [2]. TE speech is produced with the use of a TE prosthesis placed through the tracheo-esophageal wall. This allows pulmonary air to be shunted into the esophagus where it can be released through the pharyngoesophageal segment. Although TE speech is a clinically preferred method of voice rehabilitation, only a limited number of patients are rehabilitated in the long-term [3] due to reasons such as lack of tissue integrity or poor respiratory health (for reviews, see [4]–[6]). The EL is a battery-powered unit that provides a mechanical voice source through the tissues of the neck or directly into the mouth via a flexible tube. An EL is used for verbal communication in over half of laryngectomy cases [3], [7]–[9]. Although it is widely used, the EL has several limitations. Most EL devices require the dedicated use of one hand and do not provide a means to control pitch while speaking, two issues noted in the top five deficits of EL speech communication for both users and for speech-language pathologists [10]. These deficits may contribute to the lowered quality of life scores seen in electrolarynx users relative to individuals using TE speech, particularly with respect to communication ability [11].

Previously, we developed EL technology utilizing neck surface electromyography (sEMG) to control the activation, termination, and pitch of an EL, freeing both hands during speech and providing the ability to produce pitch-based intonational contrasts [12]. The EMG-controlled EL (EMG-EL) uses sEMG to provide on/off control based on a threshold. Rather than using a single sEMG envelope threshold for on/off control, the system employs an offset threshold that is an adjustable ratio of the onset threshold, creating a hysteresis band that allows the user to bias the EMG-EL toward maintaining voicing once initiated, reducing unwanted cutouts. Pitch is controlled by the level of suprathreshold sEMG energy.

The "naturalness" of speech may be assessed perceptually using discrete (numbered) scales, visual analog scales, and paired comparison ratings. Meltzner and Hillman used visual analog scale and paired comparison ratings to study the naturalness perceived among normal natural speech, normal speech with the pitch variation removed, typical EL speech, and EL speech with a natural pitch contour imposed upon it [13]. They found that the addition of a naturalistic pitch contour increased the perceived naturalness of the EL speech, whereas the subtraction of pitch variation decreased the perceived naturalness of normal speech. While the addition of a pitch contour to EL speech is assumed to improve naturalness, sEMG-controlled pitch may not have the same effect as natural pitch contours. Naturalness may also suffer if imprecision in onset and offset control during EMG-EL use decreases intelligibility.

The best EMG-EL control source would presumably come from a neural pathway normally responsible for voice production. To test this hypothesis, patients of a previous study

Manuscript received June 19, 2008; revised November 04, 2008; accepted January 13, 2009. First published March 16, 2009; current version published April 08, 2009. This work was supported in part by the National Institute of Deafness and Other Communication Disorders Grant R01-DC006449.

C. E. Stepp is with the Harvard-MIT Division of Health Sciences and Technology, Cambridge, MA 02139 USA (e-mail: cstepp@alum.mit.edu).

J. T. Heaton and R. E. Hillman are with the Department of Surgery, Harvard Medical School, Boston, MA 02115 USA (e-mail: heaton.james@mgh.harvard.edu).

R. G. Rolland was with the Institute for Health Professions, Massachusetts General Hospital, Boston, MA 02129 USA. She is now with the Learning Prep School of West Newton, MA 02465 USA (e-mail: rebecca.givens@aya.yale.edu).

Digital Object Identifier 10.1109/TNSRE.2009.2017805

had their laryngectomy surgery experimentally modified to preserve and reposition neck strap muscle bilaterally, with a natural nerve supply maintained on one side of the neck and targeted muscle reinnervation of the strap muscles by a rerouted recurrent laryngeal nerve on the contralateral side [12], [14], [15]. Both strap muscle nerve supplies were effective in the control of EMG-EL initiation, termination, and pitch modulation [14], [16], with the recurrent laryngeal nerve not providing a clear advantage for EMG-EL control on any parameter, including speech fluency and intelligibility [14], [16]. Therefore, strap muscle preservation with a natural nerve supply can provide an effective EMG-EL control source without laryngeal nerve transfer. However, these muscles are typically excised during standard total laryngectomy [17], particularly when neck dissection of lymphatics is performed, and their intentional preservation can only be considered in patients without neck disease beyond the larynx (which is often not the case with advanced laryngeal cancer requiring total laryngectomy). Given that rerouted laryngeal motor commands are not clearly advantageous for effective EMG-EL control, and that relatively few laryngectomy patients can potentially make use of residual neck strap muscles, the present study explored alternative head and neck muscle control sources typically available after standard total laryngectomy.

The primary goal of this study was to ascertain the onset and offset control capabilities of individuals who had undergone standard total laryngectomy, without special efforts to preserve musculature for EMG-EL control, using sEMG from neck and face locations to control the EMG-EL. A further goal of this study was to determine whether this untrained sEMG-control of onset, offset, and pitch resulted in perceived naturalness comparable to EL speech produced with a typical handheld EL.

II. Methods

A. Participants

Participants were eight individuals (two females, six males) with a mean age of 61 years ($R = 48$–80 years) who had undergone total laryngectomy at least one year previously and were proficient users of EL speech, even if it was not their primary mode of communication. The average time past laryngectomy was five years (R = 1–17 years). Six participants used EL speech as their primary mode of communication. Two participants were proficient users of TE speech, and had used TE speech as their primary mode of communication for at least one year, but maintained EL use for backup communication. Five of the participants had a history of radiation therapy, three pre-surgery and two post-surgery. All participants reported that they were nonsmokers during the time of the experiment, with no history of other speech, hearing, or language disorders.

B. Recording Procedure

Differential sEMG electrodes (Delsys DE2.1) consisting of two parallel bars (10 mm × 1 mm) spaced 10 mm apart were positioned at seven locations across the ventral neck and face surface, with preference to the side of the neck with the least anatomical change from surgery, based on participant information at the time of the recording. Example electrode locations

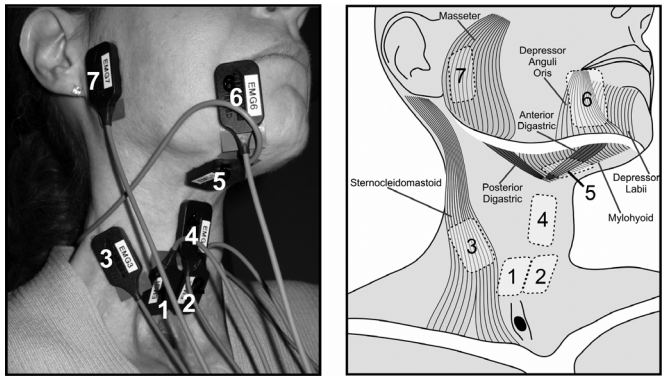

Fig. 1. Electrode placement. The left panel shows sEMG electrode locations as placed on participant S3 during experimentation. The right panel shows a schematic of expected residual muscles after total laryngectomy superficial to sEMG electrodes, depicted on the right side only. Absent from this depiction is the platysma, which is located in the subcutaneous tissue of the neck.

are shown in Fig. 1 on one participant and schematically. Electrode locations included positions 1 cm lateral to the neck midline (right and left) and just superior to the stoma (#1 and #2), centered on the sternocleidomastoid at one-third of the distance from the clavicle to the mastoid (#3), 1 cm lateral to the ventral neck midline at the superior-most location prior to the start of the submental surface (#4), 1 cm lateral to the submental midline (#5), just below the corner of the mouth (#6), and centered on the lateral jaw superficial to the masseter muscle (#7). All electrodes were referenced to a single ground electrode placed on the participant's wrist. Electrode positions #1 and #2 were placed bilaterally to record sternohyoid and sternothyroid activity, should those muscles still remain. Electrode #3 was placed in order to record from the sternocleidomastoid while avoiding known locations of muscle innervation zones e.g., [18]. Electrode #4 was placed to record from possible existing strap musculature (possibly infrahyoid and suprahyoid), while electrode #5 was intended to be sensitive to suprahyoid strap and tongue root musculature. Electrode #6 placement was intended to primarily record from depressor anguli oris and depressor labii inferioris. Electrode #7 was placed to record from the masseter muscle. When an electrode position was intended to record from a particular muscle (e.g., SCM, masseter), the electrode was aligned with its axis parallel to the supposed underlying fibers; when the target musculature was thought to be varied, electrodes were aligned such that the electrode axis was parallel to the most likely direction of most underlying fibers.

Simultaneous acoustic signals from a headset microphone (AKG Acoustics C 420 PP) and the seven channels of sEMG signals were filtered and recorded digitally (20 000 Hz sampling rate) with Axon Instruments hardware (Cyberamp 380, Digidata 1200) and software (Axoscope). An example of the audio and sEMG data collected during experimentation is shown in Fig. 2.

The participants used a commercially-available hand-held EL to produce serial speech (saying the days of the week, counting 1–10) with a pause between each word, 10 sentences randomly selected from the Yorkston and Beukelman test [19] with the instruction to leave a pause between each sentence, and a sample of spontaneous speech. Spontaneous speech samples were elicited with a question or series of questions from

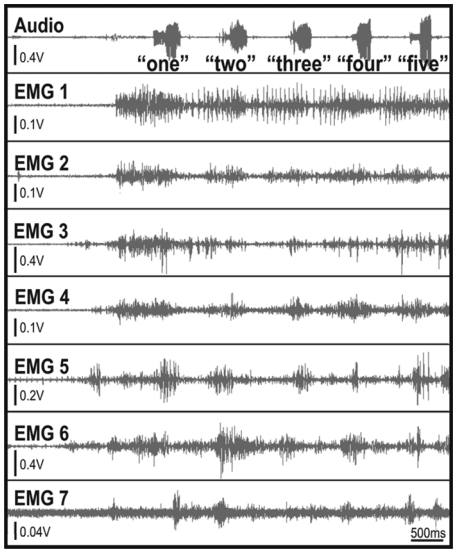

Fig. 2. sEMG electrode recordings. An example of the audio and raw sEMG data collected as a participant counted aloud using a typical EL. Traces indicated by EMG1–EMG7 refer to sEMG collected from the seven electrode positions indicated in Fig. 1.

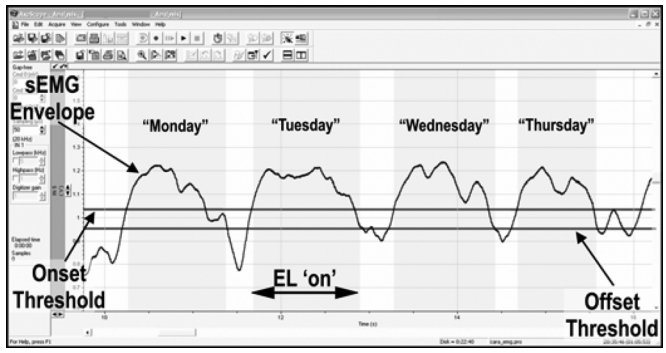

Fig. 3. Example screenshot of sEMG biofeedback. An example of the real-time visual feedback of the rms sEMG and EMG-EL threshold settings is shown. The time-varying line is the sEMG envelope used to control onset and termination of the EMG-EL. Horizontal lines specify onset and offset thresholds. Shading indicates device activation.

the investigators, typically leading to normal conversational speech (e.g., "What are your plans for this weekend?"). In the six individuals utilizing EL speech as their primary mode of communication and one of the individuals typically using TE speech, the EL used was their personal EL device. In these cases, no modifications were made to their typical settings or behavior. One individual (S3) who used TE speech as her primary mode of communication did not bring her backup EL to the recording session. This participant used a TruTone EL (Griffin Labs) from our clinical facility.

C. EMG-EL System

The EMG-EL consisted of a desktop computer running MATLAB (MathWorks), a digital signal processing board (Motorola DSP56311EVM), and an EL (NuVois). The selected sEMG signal received from the electrode being used to control the device was processed by the DSP to create a fast sEMG envelope (digital approximation of a three-pole active low pass filter with a 5 Hz 3 dB corner frequency) and a slow sEMG envelope (digital approximation of a three-pole active low pass filter with a 1 Hz 3 dB corner frequency). The fast envelope was used to control EL activation and termination whereas the slow envelope was used to modulate the EL pitch. Activation and termination thresholds were set independently for each electrode used to control the EMG-EL, with the termination threshold set at 60%–70% of the activation threshold in order to assist in uninterrupted voicing. The EMG-EL was mountable to participants using a thick, flexible copper wire bent around the base of the neck, but was hand-held in these experiments to avoid interfering with the multiple neck sEMG recording locations. Participants were provided with some basic information in preparation for using the EMG-EL. Participants were informed that stronger muscle contractions would lead to the device turning on, that relaxation would lead to the device turning off, and that increases in muscle activity would lead to increases in pitch, but they were only asked to focus specifically on device onset and offset precision.

Testing of the control capabilities began with sequential control of the EMG-EL with sEMG from each electrode recording location as the participant attempted to produce serial speech with a pause between each word. The participant was presented with real-time visual feedback of the 1-s rms sEMG and EMG-EL threshold settings for the electrode position being tested using a video monitor placed approximately 1 m away. An example screenshot of this feedback is shown in Fig. 3. After all seven electrode positions had been tested, the participant and the investigators determined the two channels they felt offered the participant the best control. For these two positions, the participant produced two types of running speech using the EMG-EL: 10 sentences selected randomly from the Yorkston and Beukelman test with the instruction to leave a pause between each sentence and spontaneous speech, elicited as described previously.

D. Perceptual Assessment of the Naturalness of EL Speech

In order to assess the naturalness of speech produced with the EMG-EL, two sets of tokens (sentences and spontaneous) were prepared for six of the eight participants in three conditions: using a typical hand-held EL and using the EMG-EL as controlled by each of the two recording locations determined at the time of the experiment to provide the best control capabilities. Two participants were excluded from the naturalness study. One of these participants used a TruTone EL with handheld pressure-based pitch control, whereas the other suffered a malfunction of the EMG-EL pitch control during the recording of the running speech. One set of tokens (referred to as sentences) was one of the ten sentences from the Yorkston and Beukelman test [19], "Some aspects of it are very interesting, but others are not." This particular token was chosen as it contained the smallest number of reading errors and other dysfluencies across all participants and modes of EL control (typical and EMG-EL choices 1 and 2). Because this speech was controlled for content, we felt it offered the best measure of control capabilities using the EMG-EL relative to a hand-held device. One sentence from the spontaneous speech sample was chosen as a token for each participant and mode of EL control to highlight the potential

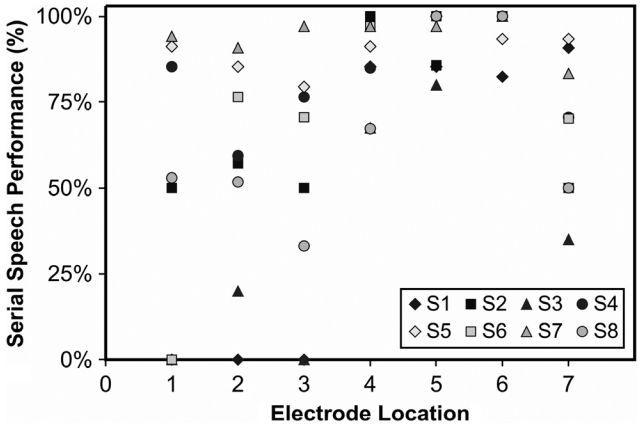

Fig. 4. Serial speech performance. The serial speech performance (average of the percentage of appropriately voiced words and the percentage of appropriately unvoiced pauses) is shown for each participant at each of the seven electrode positions.

performance abilities of the participants (referred to as spontaneous) during conversational speech. Table I shows a transcript of the spontaneous speech tokens. Tokens were chosen based on consensus of the authors on the speech sample highlighting the most natural speech produced by the participants for each condition. All of the tokens used (sentences and spontaneous) were amplitude-normalized using Adobe Audition software.

The same perceptual experimental procedure was used for both the sentences and spontaneous stimuli, consisting of visual analog scaling. Listeners were a group of 10 normal-hearing students of speech-language-pathology who had previously taken a course on voice disorders. Listeners participated in the visual analog scale experiment on two visits: one visit using the sentence stimuli and one visit using the spontaneous stimuli. On average, each listening visit took less than 1 h. The listeners were asked to compare the tokens to an anchor token of natural normal speech in which a male individual with normal voice said the sentence. Listeners could listen to this anchor at any time during the experiment. All tokens of the speech set (sentences or spontaneous) were randomly presented with a screen showing a 100 mm visual analog scale with the left end labelled "Not at all different" and the right end labelled "Very Different." The rating of each stimulus was the distance in millimeters from the left end of the scale. In order to assess intra-rater reliability, 20% of the stimuli were randomly presented a second time. The average intra-rater reliability for the visual analog scale task was evaluated using Pearson's R and was found to be 0.85 ($R = 0.58$–0.99).

After their second listening visit, listeners were asked to complete a short questionnaire concerning which aspects they thought they were attending to during the listening experiment. Listeners were asked to identify in ranked order which elements they focused on to determine naturalness and were given the following choices: prosody, intelligibility, articulation, pitch rate, content, and noise. Listeners were asked to identify in ranked order which elements they felt were most disruptive in the voices they rated negatively and were given the following choices: robotic/monotone, trouble understanding, excessive buzz, pitch too high/low, pitch moving up or down too much, rough/gravelly tone. Lastly, listeners were asked to identify which one quality the voices that were most pleasant to listen to encompassed and were given the following choices: smooth, normal-sounding pitch, melodic/lots of intonation, easy to understand, and "human-like."

TABLE I
TRANSCRIPT OF SPONTANEOUS SPEECH TOKENS

Participant	EL Control	Text (Number of Syllables)
S1	Typical	Once we went by there, we just breezed right through. (10)
	EMG 1	They say it's an amazing change, an amazing difference. (14)
	EMG 2	You've heard of tennis elbow, well now I've got research elbow. (15)
S2	Typical	And, the only thing I get when I try to do esophageal speech is a bellyache. (23)
	EMG 1	The last one we just did did not seem as good as the first time we did that one. (19)
	EMG 2	Well, I found it very interesting, sometimes a little frustrating. (18)
S4	Typical	I'm probably the only person stupid enough to get lost on the greenline. (20)
	EMG 1	Now I'll tell you about the time I boarded the wrong plane and flew to the wrong city. (12)
	EMG 2	I went to Maine this weekend. I went to Wellsby. (21)
S5	Typical	I went down to MassGeneral and went into the old jail - really nice. (15)
	EMG 1	And I would be able to tell you more. (9)
	EMG 2	It's very beautiful, really nice. (10)
S6	Typical	For somebody like me that's not an option now. Am I right? (15)
	EMG 1	So it probably sounds a lot different to you than it does to me. (17)
	EMG 2	Well, this is a pretty interesting test. (11)
S7	Typical	I grew up in the Hotel Tourraine here in Boston. It's now the Towne & Country Apartments. (20)
	EMG 1	Yeah, I'm sure it will have a lot to do with my girlfriend. (14)
	EMG 2	When I go up in the mountains for a couple of days, he's got my guns, he's got my food all packed, and I'm gone. (27)

The text of the spontaneous speech tokens used for visual analog scale testing are shown as well as the number of syllables in each token.

E. Data Analysis

Audio recordings were scored perceptually by the first author in terms of the percentage of words that were fully voiced (uninterrupted EMG-EL activation) as well as the percentage of pauses between words or sentences that were achieved. Each participant/electrode combination was scored for a percentage of achieved voicing (percentage of fully voiced words of those attempted) and of the percentage of achieved pauses during serial speech. As a simple indicator of general control ability, "serial speech performance" was defined as the average of the voicing performance (%) and pause performance (%), weighing the two equally. In a few instances, participants were completely unable to produce serial speech using the EMG-EL from particular electrode locations (S1: positions 1, 2, and 3; S3: position 1; S6: position 1), and were scored as having 0% serial speech performance.

Each of the 10 sentences was likewise scored in terms of performance, here defined as the average of the voicing performance (percentage of fully voiced words relative to the

number of words intended) and pause performance (percentage of pauses between sentences that were successfully achieved relative to the total number of intended pauses), again weighing the two equally. In addition, the sentences were also judged in a manner consistent with the perceptual scoring of sentence production using the EMG-EL in the work of Goldstein *et al.* [14]. Inasmuch, only sentences in which all words were fully voiced were counted as successful, with the final score for sentences equal to the ratio of successful sentences to the total of those attempted. To assess inter-and intra-rater reliability, 10% of the serial speech and sentence recordings were independently judged by a certified speech language pathologist and again by the first author (approximately six months later) yielding inter-rater reliability as measured with Pearson's R of 0.98 and intra-rater reliability of 0.99.

The sEMG gathered from each recording location during serial speech using a traditional EL was analyzed as the rms during the task (words and pauses) as a percent above the baseline sEMG during the participant's rest. Because the time-scales of the word and pause tasks were near 1-s (an appropriate temporal window for sEMG processing [20]), the entirety of each word or pause task was used for each rms measure, resulting in a variable temporal window. The rms sEMG was collected for words and for pauses as selected manually by the first author using visual inspection of the audio signal while listening to the audio simultaneously. Visual inspection was used to locate periods in which the EL was activated, whereas listening to the audio enabled identification of articulatory cues such that intended words and pauses could be identified. The sEMG from each electrode during words and pauses was also collected for serial speech produced using the EMG-EL while being controlled by the electrode in question. In this case, device activation was not always concurrent with the intent to speak and in some cases it was not possible to definitively identify intended words and pauses. Thus, the sEMG as a percent above baseline was only estimated for participant/electrode combinations at which the participant had achieved at least 80% perceptual "serial speech performance" to avoid the addition of error due to listener uncertainty of voicing intent. The threshold of 80% was chosen arbitrarily *a priori*. For all analysis performed, only words and pauses that were produced correctly were used; for example, if the participant failed to produce a pause between two consecutive words, neither the two words nor the absent pause would be included in analysis.

The visual analog scale data were analyzed in terms of the mean distance of each stimulus (in millimeters) from normal natural speech $(\text{score} = 0)$. A *post hoc* analysis of fundamental frequency (physical correlate of the perception of pitch) was carried out on the tokens produced with the EMG-EL and used in the visual analog scale task. This analysis used estimates of fundamental frequency calculated using Praat acoustic analysis software [21] to find the maximum, minimum, and standard deviation of fundamental frequency for each token.

III. Results

A. EMG-EL Control Performance

Serial speech performance at electrode positions #1 and #2 (inferior anterior neck), #3 (sternocleidomastoid), and #7 (mass-ester) varied greatly amongst participants, with values ranging from 0% to nearly 100% (see Fig. 4). Alternatively, positions #4 (superior neck), #5 (jaw opening muscualture), and #6 (lip depressing musculature) showed consistently high serial speech performance across all participants. These results match the subjective choices of top electrode positions made during the experiment by the participants and investigators. In all participants, at least one of the two electrode positions showing the highest serial speech performance values was one of the two electrode positions chosen during the time of the experiment; in six of the eight participants the two electrodes corresponded exactly.

TABLE II
Electrodes Chosen for Further Testing

Participant	S1	S2	S3	S4	S5	S6	S7	S8
Electrodes Tested	5, 6	4, 6	5, 6	5, 6	5, 7	5, 6	1, 5	5, 6

For each participant, the two electrodes chosen for further testing (read sentences and spontaneous speech) at the time of the experiment are listed.

Of the eight participants, five individuals had previously undergone radiation therapy. In order to assess the possible interaction between radiation therapy and the number of viable electrode recording locations, a chi-square test was performed on the serial speech performance data, assessing the counts of electrode locations showing performance values greater or equal to 80% versus performance values less than 80% in the individuals with a history of radiation therapy versus individuals with no history of radiation therapy. Again, the cutoff of 80% was chosen arbitrarily *a priori*. The results of the chi-square test showed higher than expected counts of "successful" ($\geq$80%) electrode locations in the individuals with no history of radiation therapy ($\text{df} = 1$, $p = 0.009$).

The performance during sentences was assessed at the two electrode positions chosen for further testing at the time of the experiment. Electrodes chosen for experimentation varied by participant (see Table II), although positions 4, 5, and 6 were most often chosen. Sentence production yielded consistently high results across participants for both electrode locations. The average speech performance was 97% $(\text{SD} = 2.3\%)$. When scored as in Goldstein *et al.* [14], the sentence scores were far more varied, with a range from 20% to 100% (mean $= 64\%$, $\text{SD} = 21\%$).

B. sEMG During Task Performance

During serial speech using a traditional EL as well as during the use of the EMG-EL with visual feedback, the percent above baseline rms varied significantly over both participant and electrode position. In cases in which the participant was able to achieve at least 80% serial speech performance, sEMG changes seemed to occur for both words (increase) and pauses (reduction). The difference between words and pauses in the percent above baseline rms sEMG from each recording location during serial speech using a traditional EL (labelled as "Initial") and during use of the EMG-EL with visual feedback (labelled as "During Feedback") is shown in Fig. 5. Specifically, in an effort to produce voiced serial speech with pauses between words, participants generally attempted to increase the sEMG during words, decrease the sEMG during pauses, or a combination of

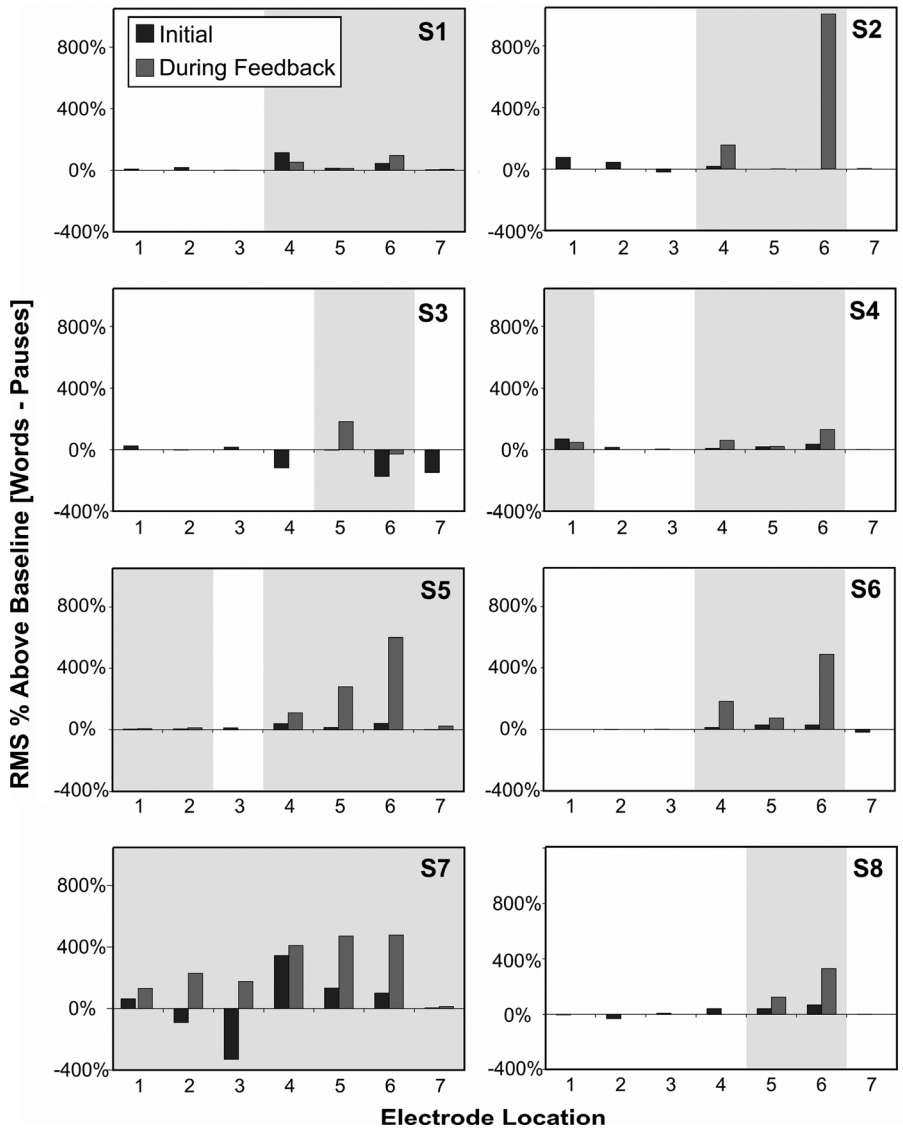

Fig. 5. Participant sEMG during serial speech. The difference between words and pauses in the percent above baseline rms sEMG from each recording location during serial speech using a traditional EL ("Initial") and during use of the EMG-EL with visual feedback ("During Feedback"). Background shading indicates electrode positions for which each participant achieved at least 80% serial speech performance using the EMG-EL.

the two approaches. A t-test comparing the percent above baseline rms during word production initially to that during feedback found a statistically significant increase (mean $= 108\%$, $p = 0.03$, one-way paired t-test, $\mathrm{df} = 56$). A t-test comparing the percent above baseline rms during pause production initially to that during feedback found a nonsignificant decrease of 37% (one-way paired t-test, $p = 0.20$, $\mathrm{df} = 56$). Moreover, a comparison of the difference between the percent above baseline rms during words and pauses showed a significant increase during feedback relative to the initial condition (mean $= 145\%$, one-way paired t-test, $p < 0.0001$, $\mathrm{df} = 56$).

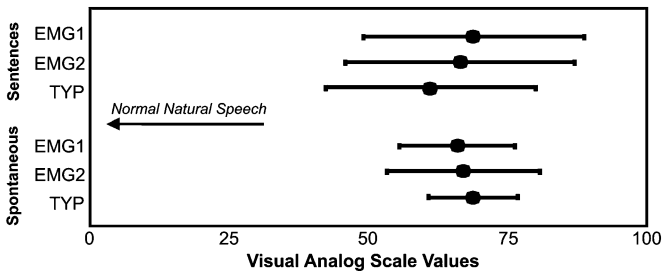

Fig. 6. Visual analog scale perceptual results. The results of the visual analog scale assessment are shown graphically, separated by speech task ("Sentences" and "Spontaneous"). Error bars extend $+/-$ one standard deviation. Normal natural speech is located at 0.

C. Naturalness

Average results of the visual analog scale assessment are shown graphically in Fig. 6. A four-factor ANOVA was performed on the dependent variable of the visual analog scale scores. The four factors were speaker, listener, EL mode (typical versus EMG-EL choices 1 and 2), and speech type (sentences versus spontaneous). The ANOVA showed that, although there was a general trend for higher visual analog scale values (less natural) for spontaneous speech than sentences, there was not a significant main effect ($p = 0.08$). Likewise, the mode of

EL control (typical versus EMG-EL choices 1 and 2) did not produce a significant main effect ($p = 0.68$). In fact, when the data are examined separately by speech type (sentences versus spontaneous), the trend of EL control source reverses (see Fig. 6) corroborating the idea that there were not significant effects as a result of EL control source. The ANOVA did show a highly significant main effect for both speaker and listener ($p < 0.001$). Interestingly, when the main effect was examined as a function of speaker, the individual with the most natural average visual analog scale value (S1, average $= 50.0$) was a retired professional radio and television announcer, whereas the individual with the least natural average visual analog scale value (S5, average $= 86.4$) was an individual who had undergone extensive tongue-base dissection during laryngectomy, resulting in some articulatory issues. Three of the listeners used were found to have intra-rater reliability (Pearson's R) less than 0.80 (0.58, 0.61, 0.75). The perceptual data were reanalyzed as above with the judgments of these three listeners excluded. Again, the four-factor ANOVA only showed significant differences for speaker and listener ($p \leq 0.001$).

Although no significant difference was found in the perceived naturalness of the sentence and spontaneous speech tokens, a *post hoc* analysis of the fundamental frequency content of the tokens was performed to determine what role, if any, the fundamental frequency contour may have played in perceived naturalness. Examination of the fundamental frequency range (maximum–minimum) and fundamental frequency standard deviation in the sentence and spontaneous tokens produced with the EMG-EL showed no difference between the two types of speech tokens for either parameter based on a two-sample unpaired *t*-test ($p = 0.91, \mathrm{df} = 21$; $p = 0.96, \mathrm{df} = 21$). The average fundamental frequency range was found to be 20.8 and 21.4 Hz for the read sentence and spontaneous tokens, respectively. The average fundamental frequency standard deviation for both speech types found was 4.1 Hz. Neither parameter showed a significant correlation with average visual analog scale values ($p = 0.76$ and $p = 0.96$, respectively). Despite the ability to provide fundamental frequency control using the EMG-EL, fundamental frequency content was not optimized in this study, and the fundamental frequency content of these EMG-EL speech materials was not comparable to the normal natural speech used for comparison. The token used as an example of normal natural speech had fundamental frequency content with a range of 79.4 Hz and standard deviation of 16.9 Hz, well above the EMG-EL tokens.

Listener questionnaire results indicated that when the listeners were asked to rank elements on which they focused to determine naturalness, the element most often in the top two choices was prosody (chosen by eight of ten listeners), followed by articulation (chosen by four), and intelligibility (chosen by three). When listeners were asked to rank which elements they felt were most disruptive in the voices they rated negatively, the elements most commonly in the top two choices were robotic/monotone (chosen by eight) and trouble understanding (chosen by eight). When listeners were asked to identify which one quality the voices that were most pleasant to listen to encompassed, "melodic/lots of intonation" was chosen by seven listeners, "easy to understand" by two listeners, and "human-like" by one listener.

IV. Discussion

A. Physiological Correlates of Task Performance

All participants showed high serial speech performance when using sEMG from electrode recording locations #4, #5, and #6. Vocal-related activity from these recording locations likely stemmed from residual suprahyoid and tongue root musculature (used for articulation and laryngeal control in healthy individuals) and possibly platysma (#4 and #5) as well as the depressor anguli oris (#6). For all participants, the superior ventral neck or submental surface (#4 or #5) was at least one of their two best control locations, leading to average serial speech performance of 95% (SD = 8%). The face recording location below the corner of the mouth (superficial to the depressor anguli oris; #6) was also an effective control location for all participants, but presents an unfavorably conspicuous electrode site and may be more prone to false triggering with non-speech lip movements.

Two likely reasons for poor performance using a recording site would be the result of a lack of natural correspondence between speech and sEMG from the site and/or loss of relevant tissue in some participants due to surgical intervention and/or radiation therapy. The tissue integrity at site #7 seems unlikely to be affected by most total laryngectomy surgeries or radiation therapy; inconsistency in the performance at this site may be related to differences in articulation and/or electrode placement across participants. The inconsistency in performance in electrodes #1 and #2 is likely to be a result of the loss of relevant tissue in some participants due to surgical intervention and/or radiation therapy. While it is possible that participants experienced a loss of tissue integrity near electrode #3 (sternocleidomastoid), a more likely explanation for variable performance at this location is a lack of natural correspondence between muscle activity and speech production. The sternocleidomastoid may be utilized for complex breathing and singing performance, but is thought to be largely inactive for "simplified speaking tasks" such as those attempted here [22].

The right panel of Fig. 1 shows a schematic of the electrode recording locations relative to the musculature thought to be left after total laryngectomy. Absent from this depiction is the platysma, which is a superficially located thin sheet of muscle in the subcutaneous tissue of the neck. It extends over the anterolateral aspect of the neck from the inferior border of the mandible to the superior aspect of the pectoralis major. During total laryngectomy, the neck is incised near the clavicle and an apron-shaped skin flap is raised toward the head, preserving most of the platysma length attached to the lower jaw, and maintaining its motor supply through the cervical branch of the facial nerve [17]. The likely retraction of deeper neck muscles (e.g., strap muscles) after their division during laryngectomy and the probable survival and superficial location of the platysma makes it a potential source for the sEMG collected at electrode recording sites #1, #2, #3, #4, and possibly #5 in this study. Although the activation of the platysma during speech has been studied less than other laryngeal and orofacial musculature, it has been shown to be active during speech production [23]. It is thought to be an antagonist to the orbicularis oris inferior muscle, and has been shown to activate during lowering of the lower lip during speech and nonspeech tasks [23]. It is

possibly active during a large selection of phonemes created by lip movements. Therefore, platysma-based sEMG could provide consistent EMG-EL control during running speech, but would perhaps perform weakly for control of the EMG-EL for nonarticulated speech such as prolonged vowels produced without lip rounding.

B. Possible Effects of Radiation Therapy

A history of radiation therapy may be indicative of a loss of muscle integrity. Comparison of overall serial speech performance between individuals with a history of radiation therapy and those without showed that those individuals who had had radiation tended to have fewer electrode recording positions yielding at least 80% serial speech performance (chi-square test, $\mathrm{df} = 1$, $p = 0.009$); this finding also held up when limiting the data used for testing to the fraction of the data collected at sites 1–5 (neglecting #6 and #7), which are the sites most likely to be affected by radiation therapy (chi-square test, $\mathrm{df} = 1$, $p = 0.022$). Perhaps the radiation therapy reduced the muscle integrity, leading to reduced overall control capabilities; however, it is also possible that the need for radiation therapy covaries with the need for more extensive surgical intervention. Therefore, we cannot determine that there is a causative relationship between a history of radiation therapy and the ability of an individual to control the EMG-EL.

C. Comparison of Sentence Performance With Previous Findings

Using their two best electrode recording locations for EMG-EL control, all participants were able to produce running speech (sentences) with few disrupted words due to breaks in voicing. This replicates the findings of Goldstein and colleagues who found that individuals controlling the EMG-EL with neck strap muscles showed improvement in their ability to produce sentences without formal training, and that running speech may be more easily produced with the EMG-EL due to the ability of participants to anticipate pauses and adjust muscle activity accordingly [14]. When sentence production was scored similarly to Goldstein *et al.* [14], the average score for all participants at both electrode sites tested was 64% ($\mathrm{SD} = 21\%$). The three individuals with laryngectomy studied by Goldstein *et al.* had an average score of 30% ($\mathrm{SD} = 26\%$) prior to training and 83% ($\mathrm{SD} = 21\%$) after training [14] using neck strap muscle surgically modified to be innervated by the RLN. Their training consisted of 4–10, 10–60 min training sessions. Without any formal training or surgical modifications, the individuals with laryngectomy studied here were able to produce sentences at least as well as the individuals in Goldstein *et al.* before training. Future work will include a study on the effects of training on our present participants' ability to control the EMG-EL. Mimicking the training protocol and outcome measures employed in studies of individuals controlling the EMG-EL with RLN-innervated neck strap muscle will reveal whether EMG-EL control capabilities are enhanced by modification of the laryngectomy surgery.

The sentence scoring scheme of Goldstein *et al.* resulted in drastically different scores than the one utilized here. Their method implied the stringent rule that only sentences in which all attempted words were fully voiced were counted as successful. While this rigorous approach is ideal for discriminating small differences in high performing individuals, it does not relate well to assessment of overall communicative ability. Any brief interruption in one word of a sentence would result in that sentence being designated with a "0" score, although the communicative intent of the speaker was likely completely understood. We believe our approach here corresponds more closely with the actual ability of these individuals to communicate with the EMG-EL.

D. Effects of Biofeedback

The fact that the difference between the percent above baseline rms sEMG during words and pauses showed a highly significant increase during feedback relative to the initial condition suggests that while participants may have employed differing strategies (increasing sEMG during words or decreasing sEMG during pauses), the chief result was an increase in the dynamic range between the sEMG during words versus pauses. Further, these statistically significant changes suggest that attendance to relevant muscle groups using visual sEMG feedback can improve EMG-EL control, without the use of formal training protocols.

E. Naturalness of EL Speech

Based on the previous work of Meltzner and Hillman [13], using the EMG-EL might provide an advantage in perceived naturalness relative to typical EL speech due to the added fundamental frequency fluctuations. However, the perceptual testing employed here failed to show a significant difference in the perceived naturalness of both read speech and spontaneous speech using a typical handheld EL versus the EMG-EL. This is likely the result of two factors: differing on/off control and underutilization of fundamental frequency capabilities. Participants using the typical EL device were experienced and proficient at utilizing the button switch to produce precise on/off control. Conversely, participants using the EMG-EL to produce speech had little experience in controlling an EL with neck and face musculature, producing continuous speech with occasional cutouts and unintended voice prolongations. These cutouts were likely detrimental to listener judgments of naturalness. The fundamental frequency control of the EMG-EL is set to be linearly related to the level of suprathreshold sEMG energy, with the related coefficient set by the user (or the experimenters in this case). The results of the *post hoc* fundamental frequency analysis showed that the fundamental frequency settings employed for the EMG-EL users provided fundamental frequency fluctuations with range and standard deviation far lower than those seen in normal natural speech. Future work will incorporate recordings with optimized fundamental frequency control settings to fully explore the possible gains in naturalness.

F. Comparison of the EMG-EL With Other Alaryngeal Rehabilitation Methods

A future version of the EMG-EL may provide a better option for alaryngeal speech rehabilitation for individuals with laryngectomy in terms of the ability to easily control the device without the use of a hand and the ability to intuitively modulate fundamental frequency, although it currently suffers with respect to alternative devices in terms of usability.

With regard to hand-use, the traditional EL does not provide hands-free control, whereas some TE prostheses and newer EL devices may be used hands-free. A TE prosthesis may be fit with a stomal valve such that higher pressures (distinct from normal expiration) force air from the lungs through the prosthesis, creating TE voicing without requiring hand closure of the stoma. Esophageal speech is by nature hands-free, but is difficult to acquire (for review, see [24]), requires high motivation and intensive speech therapy, and has declined in popularity since the advent of the TE prostheses in the early 1980s. For EL users, Griffin Laboratories (Temecula, CA) has developed a holder for two of their EL models that allows for hands-free control based on depression of the chin against the neck-mounted EL. In addition to requiring somewhat unnatural head movements for EL activation, this apparatus requires that the user's neck tissues adequately transmit sounds originating from a relatively high location along the neck midline, which is not possible for many laryngectomees. The EMG-EL also provides on/off control, but it is based on a more vocal-related physiological control source. However, false triggering due to muscle activity that is not intended for speech decreases the reliability of EMG-EL on/off control. In the initial (analog) prototype of this device, it was outfitted with a momentary mute button to help mitigate false-triggering during swallows, coughing, and etcetera. Although this improves the reliability of EMG-EL control, further adjustments should be made in order for it to provide hands-free on/off control as reliably as that provided by TE speech and the neck-mountable EL devices.

Most common EL models are monotonic and do not provide dynamic pitch modulation, which has been identified as a principle cause for the unnatural nature of EL speech [13]. However, the TruTone EL (Griffin Laboratories) provides the ability to modulate fundamental frequency based on its pressure-sensitive on/off button; higher pressure on the button corresponds to higher pitches. This capability enables natural-sounding pitch contours for some skilled users, but can be difficult to use by others, resulting in unnatural and rapid changes in pitch. The pressure-sensitive on/off button control for pitch had been a feature on some of the earliest ELs in the late 1950s [10], [25], but difficulty in mastering hand-controlled pitch is likely why this feature was eliminated from most subsequent EL designs. With an appropriate muscular control source and training, pitch modulation generated by the EMG-EL could potentially be more natural than that created with manual (hand) control, due to the more natural correspondence between increases in neck and face muscle activation and increased vocal amplitude and pitch.

Both TE and esophageal speech provide the ability to modulate pitch based on increases in the air pressure used to drive the tissues, but pitch range is substantially restricted compared to laryngeal phonation, and habitual pitch is inappropriately low ($<$100 Hz) for these alaryngeal voice methods—particularly for females, where their habitual fundamental frequencies are normally 1–2 octaves higher. The fundamental frequency and range of pitch modulation of the EMG-EL can be user-defined through settings of the DSP unit. Further study is needed to determine the degree to which natural-sounding pitch modulation can be acquired through a combination of training and appropriate setting of the pitch modulation capabilities of the EMG-EL.

Despite the need for manual operation and the unnatural, monotone voice of traditional ELs, they offer high usability. They are small, inexpensive, lightweight, easy to learn how to use for most individuals, and can be used almost immediately post-surgery (particularly with an oral adapter). TE and esophageal speech offer the greatest level of usability for individuals who are able to adopt these techniques, producing speech that is generally more intelligible and has better vocal quality than traditional electrolarynx speech [26]. Esophageal speech requires no device to produce, but is very difficult to master, and has conspicuously limited phrase length (averaging 3.0 words per phrase relative to laryngeal speakers at 9.8 words per phrase) [27]. TE speech requires surgical intervention for valve placement (although this is often done at the time of laryngectomy), and requires persistent hygienic care of the prosthesis to avoid bacterial and fungal growth that can make it fail or become a health risk [6]. Moreover, successful TE valve placement and maintenance requires tissue integrity of the tracheal and adjacent pharyngoesophageal walls to chronically hold the valve in place, in addition to good respiratory health, which is needed to drive air through the valve to vibrate the pharyngoesophageal tissues for voice production [6]. Therefore, motivational, anatomical, and physiological constraints limit the laryngectomee population in which TE speech is a viable option.

The EMG-EL at present does not offer greater usability than any of the currently available rehabilitation modes. The first commercially available EMG-EL model would likely be smaller than current prototypes, but would still involve more components (i.e., DSP unit, EMG electrode, and EL transducer) than alternative alaryngeal voice sources, and require careful electrode placement. The system might also be relatively expensive if it is not covered by medical insurance or by special programs through state governments and telephone companies (stemming from the Americans with Disabilities Act). However, based on the fact that the present individuals were able to use the EMG-EL without formal training to produce running speech, and that their speech was perceived as natural as that produced with a typical handheld EL with which they have had significant previous experience, use of the EMG-EL in this population is promising. Future studies will assess whether increased training and manipulations of EMG-EL settings can provide users with voice control that is superior to other alaryngeal voicing sources. If so, this would justify further development of EMG-EL technology, increasing usability.

V. Conclusion

All participants were able to produce running and serial speech hands-free with the EMG-EL controlled by sEMG from

multiple recording locations, with the superior ventral neck or submental surface locations providing at least one of the two best control locations. Vocal-related activity from these recording locations likely stemmed from residual suprahyoid and tongue root musculature, and possibly the platysma. The face recording location below the corner of the mouth (superficial to the depressor anguli oris) was also an effective control location for all participants, but presents an unfavorably conspicuous electrode site and may be more prone to false triggering with lip movements. Without formal training, each participant had multiple sEMG recording locations providing intuitive and effective prosthetic voice control perceived as natural as a typical handheld EL without formal training, indicating promise for use of an EMG-EL system across a large segment of the laryngectomy population.

ACKNOWLEDGMENT

The authors would like to thank Dr. J. Kobler and Dr. J. Perkell for their helpful comments, and also the participants of this study, who gave graciously of their time.

REFERENCES

[1] *Cancer Facts & Figures 2008*. Atlanta, GA: Am. Cancer Soc., 2008.

[2] G. A. Gates, W. Ryan, J. C. Cooper, Jr., G. F. Lawlis, E. Cantu, T. Hayashi, E. Lauder, R. W. Welch, and E. Hearne, "Current status of laryngectomee rehabilitation: I. Results of therapy," *Am. J. Otolaryngol.*, vol. 3, pp. 1–7, 1982.

[3] W. M. Mendenhall, C. G. Morris, S. P. Stringer, R. J. Amdur, R. W. Hinerman, D. B. Villaret, and K. T. Robbins, "Voice rehabilitation after total laryngectomy and postoperative radiation therapy," *J. Clin. Oncol.*, vol. 20, pp. 2500–2505, 2002.

[4] C. Gress and M. Singer, "Tracheoesophageal voice restoration," in *Contemporary Consideration in the Treatment and Rehabilitation of Head and Neck Cancer: Voice, Speech, and Swallowing*, P. Doyle and R. L. Keith, Eds. Austin, TX: Pro-Ed, 2005, pp. 431–452.

[5] A. M. Pou, "Tracheoesophageal voice restoration with total laryngectomy," *Otolaryngol. Clin. North Am.*, vol. 37, pp. 531–545, 2004.

[6] G. Monahan, "Clinical Troubleshooting with Tracheoesophageal puncture voice prostheses," in *Contemporary Considerations in the treatment and Rehabilitation of Head and Neck Cancer: Voice, Speech, and Swallowing*, P. Doyle and R. L. Keith, Eds. Austin, TX: Pro-Ed, 2005, pp. 481–502.

[7] S. Gray and H. R. Konrad, "Laryngectomy: Postsurgical rehabilitation of communication," *Arch. Phys. Med. Rehabil.*, vol. 57, pp. 140–142, 1976.

[8] R. E. Hillman, M. J. Walsh, G. T. Wolf, S. G. Fisher, and W. K. Hong, "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," *Ann. Otol. Rhinol. Laryngol. Suppl.*, vol. 172, pp. 1–27, 1998.

[9] H. L. Morris, A. E. Smith, D. R. Van Demark, and M. D. Maves, "Communication status following laryngectomy: The Iowa experience 1984–1987," *Ann. Otol. Rhinol. Laryngol.*, vol. 101, pp. 503–510, 1992.

[10] G. S. Meltzner, R. E. Hillman, J. T. Heaton, K. M. Houston, J. B. Kobler, and Y. Qi, "Electrolaryngeal speech: The state of the art and future directions for development," in *Contemporary Considerations in the Treatment and Rehabilitation of Head and Neck Cancer: Voice, Speech, and Swallowing*, P. C. Doyle and R. L. Keith, Eds. Austin, TX: Pro-Ed, 2005, pp. 571–590.

[11] C. Finizia and B. Bergman, "Health-related quality of life in patients with laryngeal cancer: A post-treatment comparison of different modes of communication," *Laryngoscope*, vol. 111, pp. 918–923, 2001.

[12] E. A. Goldstein, J. T. Heaton, J. B. Kobler, G. B. Stanley, and R. E. Hillman, "Design and implementation of a hands-free electrolarynx device controlled by neck strap muscle electromyographic activity," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 2, pp. 325–332, Feb. 2004.

[13] G. S. Meltzner and R. E. Hillman, "Impact of aberrant acoustic properties on the perception of sound quality in electrolarynx speech," *J. Speech Language Hear. Res.*, vol. 48, pp. 766–779, 2005.

[14] E. A. Goldstein, J. T. Heaton, C. E. Stepp, and R. E. Hillman, "Training effects on speech production using a hands-free electromyographically controlled electrolarynx," *J. Speech Language Hear. Res.*, vol. 50, pp. 335–351, 2007.

[15] J. T. Heaton, E. A. Goldstein, J. B. Kobler, S. M. Zeitels, G. W. Randolph, M. J. Walsh, J. E. Gooey, and R. E. Hillman, "Surface electromyographic activity in total laryngectomy patients following laryngeal nerve transfer to neck strap muscles," *Ann. Otol. Rhinol. Laryngol.*, vol. 113, pp. 754–764, 2004.

[16] H. Kubert, C. E. Stepp, S. M. Zeitels, M. Walsh, S. R. Prakash, R. E. Hillman, and J. T. Heaton, "Electromyographic control of a hands-free electrolarynx using neck strap muscles," *J. Commun. Disorders*.

[17] F. Wong, "Total Laryngectomy," in *Atlas of Head & Neck Surgery—Otolaryngology*, B. Bailey, K. Calhoun, A. Coffey, and J. G. Neely, Eds. Philadelphia, PA: Lippincott-Raven, 1996, p. 934.

[18] D. Falla, P. Dall'Alba, A. Rainoldi, R. Merletti, and G. Jull, "Location of innervation zones of sternocleidomastoid and scalene muscles—A basis for clinical and research electromyography applications," *Clin. Neurophysiol.*, vol. 113, pp. 57–63, 2002.

[19] K. M. Yorkston and D. R. Beukelman, "Communication efficiency of dysarthric speakers as measured by sentence intelligibility and speaking rate," *J. Speech Hear. Disorders*, vol. 46, pp. 296–301, 1981.

[20] H. J. Hermens, B. Freriks, R. Merletti, D. F. Stegeman, J. Blok, R. Gunter, C. Disselhorst-Klug, and G. Hagg, *European Recommendations for Surface ElectroMyoGraphy: Results of the SENIAM Project*. Enschede, The Netherlands: Roessingh Research Development, 1999.

[21] P. Boersma and D. Weenink, Praat: Doing phonetics by computer 5.0.20 ed [Online]. Available: http://www.praat.org/, 2008

[22] V. Pettersen, K. Bjorkoy, H. Torp, and R. H. Westgaard, "Neck and shoulder muscle activity and thorax movement in singing and speaking tasks with variation in vocal loudness and pitch," *J. Voice*, vol. 19, pp. 623–634, 2005.

[23] M. D. McClean and S. Sapir, "Surface electrode recording of platysma single motor units during speech," *J. Phonetics*, vol. 8, pp. 169–173, 1980.

[24] P. Doyle and T. L. Eadie, "Pharyngoesophageal segment function: A review and reconsideration," in *Contemporary Considerations in the Treatment and Rehabilitation of Head and Neck Cancer: Voice, Speech, and Swallowing*, P. Doyle and R. L. Keith, Eds. Austin, TX: Pro-Ed, 2005, pp. 521–544.

[25] R. L. Keith, J. C. Shanks, and P. C. Doyle, "Historical highlights: Laryngectomy rehabilitation," in *Contemporary Considerations in the Treatment and Rehabilitation of Head and Neck Cancer: Voice, Speech, and Swallowing*, P. C. Doyle and R. L. Keith, Eds. Austin, TX: Pro-Ed, 2005, pp. 17–58.

[26] S. E. Williams and J. B. Watson, "Speaking proficiency variations according to method of alaryngeal voicing," *Laryngoscope*, vol. 97, pp. 737–739, 1987.

[27] J. Robbins, "Acoustic differentiation of laryngeal, esophageal, and tracheoesophageal speech," *J. Speech Hear. Res.*, vol. 27, pp. 577–585, 1984.

Authors' photographs and biographies not available at the time of publication.