Electromyographic control of a hands-free electrolarynx using neck strap muscles

Heather L. Kubert [a,b], Cara E. Stepp [a,c], Steven M. Zeitels [c,d], John E. Gooey [e,f], Michael J. Walsh [f], S.R. Prakash [a,c], Robert E. Hillman [a,b,c,d], James T. Heaton [a,b,d,*]

[a] *Massachusetts General Hospital Center for Laryngeal Surgery and Voice Rehabilitation, Boston, MA, United States*
[b] *Massachusetts General Hospital Institute for Health Professions, Boston, MA, United States*
[c] *Harvard-MIT Division of Health Science & Technology, Cambridge, MA, United States*
[d] *Department of Surgery, Harvard Medical School, Boston, MA, United States*
[e] *VA Health Center Boston, Boston, MA, United States*
[f] *Department of Otolaryngology/Head and Neck Surgery, Boston University Medical Center Hospital, Boston, MA, United States*

Abstract

Three individuals with total laryngectomy were studied for their ability to control a hands-free electrolarynx (EL) using neck surface electromyography (EMG) for on/off and pitch modulation. The laryngectomy surgery of participants was modified to preserve neck strap musculature for EMG-based EL control (EMG-EL), with muscles on one side maintaining natural innervation and those on the other side receiving a transferred recurrent laryngeal nerve (RLN). EMG from each side of the neck controlled the EMG-EL across a day of unstructured practice followed by a day of formal training, including EMG biofeedback. Using either control source, participants spoke intelligibly and fluently with the EMG-EL before formal training. This good initial performance did not consistently improve across testing for either control source in terms of voice timing, speech intelligibility, fluency, and intonation of interrogative versus declarative sentences. Neck strap muscles have activation patterns capable of simple alaryngeal voice control without requiring RLN transfer.

Learning outcomes: The reader will better understand (1) functionality of the hands-free electrolarynx (2) modification of laryngectomy surgery to preserve neck strap musculature and (3) performance of hands-free electrolarynx with different control sources.

1. Introduction

Approximately 12,250 new cases of laryngeal cancer will be diagnosed in the United States in 2008 (Cancer facts & figures, 2008). A subset of these cases will be treated with radical surgical intervention, including total laryngectomy.

* Corresponding author at: MGH Voice Center, 1 Bowdoin Sq., 11th Floor, Boston, MA 02114, United States. Tel.: +1 617 726 0211; fax: +1 617 726 0222.
 E-mail addresses: hkubert@hotmail.com (H.L. Kubert), cstepp@mit.edu (C.E. Stepp), zeitels.steven@mgh.harvard.edu (S.M. Zeitels), John.Gooey@med.va.gov (J.E. Gooey), Mike.Walsh@bmc.org (M.J. Walsh), srp@mit.edu (S.R. Prakash), hillman.robert@mgh.harvard.edu (R.E. Hillman), James.Heaton@mgh.harvard.edu (J.T. Heaton).

With the larynx removed, the natural sound source for speech production is also lost. Many individuals are able to produce alaryngeal voice after laryngectomy by directing air through a one-way valve that is surgically implanted in the back of the tracheal wall via tracheoesophageal puncture (TEP) to vibrate the tissues of the upper esophagus and pharynx. However, TEP is either not performed, or is attempted and subsequently removed in many laryngectomees due to difficulties in generating this type of alaryngeal voice and/or properly maintaining their air valve (Monahan, 2005). Because of these precluding factors, and the ease with which a serviceable voice is typically achieved using an electrolarynx (EL), more than half of laryngectomees use an EL as their primary means of verbal communication (Gray & Konrad, 1976; Hillman, Walsh, Wolf, Fisher, & Hong, 1998; Mendenhall et al., 2002; Morris, Smith, Van Demark, & Maves, 1992).

Notwithstanding its widespread use, the EL has multiple drawbacks that detract from the verbal communication of the user, particularly when therapeutic protocols for training effective use of EL communication (Doyle, 2005) have not been applied. Specifically, most EL devices require the dedicated use of one hand to function and typically do not provide a means to control pitch while speaking. These two issues were noted in the top five deficits of EL speech communication for both users and Speech-Language Pathologists (Meltzner et al., 2005). An exception is the TruTone® EL by Griffin Laboratories, which offers dynamic pitch modulation via a pressure-sensitive activation button and hands-free operation using a neck-mounting accessory. Unfortunately, the hands-free mounting system places the EL transducer at a sub-optimal chin/neck location for many users, and mastery of pitch modulation by varying activation pressures can be difficult for some laryngectomees, resulting in awkward vocal inflection and frustration to the point of preferring a monotone EL model.

We have previously developed new EL technology that utilizes neck surface electromyographic (EMG) signals from preserved neck strap muscles 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 (Goldstein, Heaton, Kobler, Stanley, & Hillman, 2004). The EMG-controlled EL (EMG-EL) uses the EMG recorded at the neck surface to provide on/off control. The device is activated when the envelope of the EMG signals are higher than a pre-determined threshold. Pitch of the EMG-EL is controlled by the level of suprathreshold EMG energy, with greater EMG energy corresponding to higher fundamental frequency (F0).

Previously, a cohort of participants were studied who received an experimental modification to their total laryngectomy surgery, involving targeted muscle reinnervation (TMR) by rerouting the recurrent laryngeal nerve (RLN) on one side of the neck into a set of host strap muscles through the distal trunk of the ansa cervicalis nerve. In some participants, naturally innervated (ansa cervicalis) strap muscles (ANSA-straps) were also preserved on the contralateral side of the neck as a point of comparison (Goldstein et al., 2004; Goldstein, Heaton, Stepp, & Hillman, 2007; Heaton et al., 2004). The EMG signals from both sets of strap muscles were recorded periodically for more than 1-year post-surgery in a cohort of 10 participants. Recordings occurred for each participant a minimum of once within 1–5 months after laryngectomy (early post-surgical period) and again at 13 months or greater after laryngectomy (late post-surgical period), with recordings occurring an average of every 2.7 months during the first 13 months. All participants demonstrated a conspicuous increase in EMG activity from RLN-innervated strap muscles across the first post-surgical year, with a statistically significant increase in EMG activity (RMS % increase above baseline) for the early versus late post-surgical period as a group. Moreover, the RLN-innervated strap muscles in the late post-surgical period were, on average, larger in EMG magnitude and more correlated with phonation than the naturally innervated strap muscles, suggesting that RLN transfer may provide better control capabilities for the EMG-EL (Heaton et al., 2004). Further, we have shown that training is effective in improving the control of EMG-EL activation, termination, and pitch modulation using EMG from RLN-innervated strap muscles (Goldstein et al., 2007).

This prospective study of three total laryngectomy patients extends the original cohort of individuals receiving TMR by making a direct comparison of EMG-EL control capabilities of RLN-innervated strap muscles to those of the naturally innervated strap muscles within the same individuals. Participants used the EMG-EL with each control source for a day without training (beyond basic instruction) and for a day with visual EMG biofeedback-based training. The onset, duration, termination, intelligibility, fluency, and pitch modulation capabilities of EMG-EL speech were systematically examined using both their naturally innervated strap muscles as well as their RLN-innervated muscles as EMG-EL control sources. Outcome measures were gathered before, throughout, and after biofeedback-based training.

2. Materials and methods

2.1. Participants

Participants in this experiment were three adult males at age 63, 65 and 41 (S1, S2, and S3, respectively) who had undergone a modified total laryngectomy surgery at least 1 year previously. The modified surgery with TMR involved transferring one RLN to denervated neck strap muscles on the ipsilateral side of the neck (sternohyoid, sternothyroid, and omohyoid), while maintaining naturally innervated strap muscles on the contralateral side of the neck. The details of this procedure are described elsewhere (Heaton et al., 2004). One of the three participants had a tracheoesophageal (TE) prosthesis and was a proficient user of both TE and EL speech. The other two participants exclusively used an EL for daily, proficient communication. The three participants had undergone neck surface EMG recordings an average of every 2.3 months during their first post-surgical year, and consistent EMG signals were recorded from each participant from both the naturally innervated and the RLN-transferred sides of the neck by the time EMG-EL training was initiated at a year or greater after surgery.

2.2. Instrumentation

Participants were seated in a sound-attenuating chamber, with two video monitors (22 in. LCDs) placed approximately 1 m away. One video monitor presented stimulus materials and one presented visual EMG feedback and EMG threshold settings (Fig. 1). The EMG was obtained from two differential skin surface electrodes (DE2.1; DelSys Inc., Boston, MA) placed on the neck surface superficial to neck strap muscles. One EMG electrode was positioned lateral to the neck midline, superficial to strap muscles receiving the transferred RLN nerve supply, while the other was located similarly on the contralateral side of the neck, superficial to naturally innervated strap muscles. The EMG-EL was controlled by only one electrode's signal at any time. An Ag/AgCl gel ground electrode was placed on the shoulder.

The EMG-EL system consisted of a desktop computer running MATLAB (MathWorks, Natick, MA), an analog circuit for producing an EMG envelope, a digital signal processing (DSP) board (DSP56311EVM, Motorola, Schaumburg, IL), and an EL (NuVois, Mountain Precision Mfg., Boise, ID). The EMG signals received from the
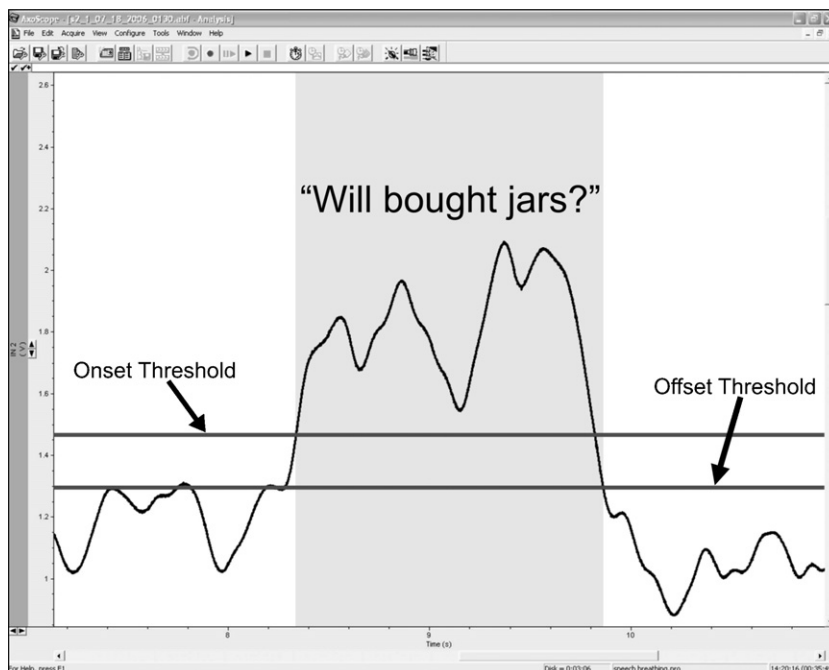

Fig. 1. Computer screen view provided for biofeedback of EMG envelope in relation to EMG-EL voice activation and termination. One horizontal bar represented the activation threshold and a second bar represented the termination threshold, with vertical shading present when the EMG-EL was producing voice.
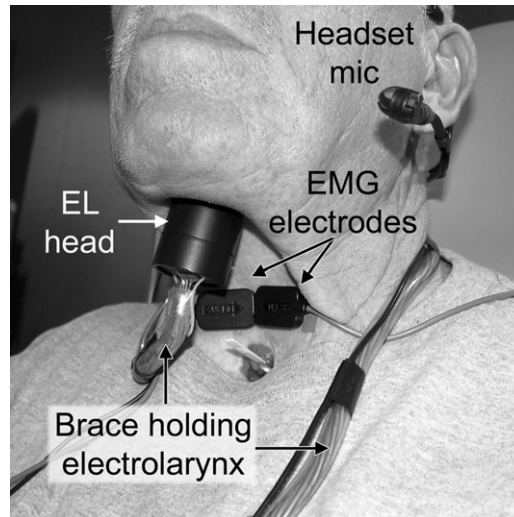
Fig. 2. The EMG-EL is shown on an individual. The EL head (transducer) and EMG electrodes are worn on the neck for hands-free EL speech.

electrode were sent to the analog circuit for creating an EMG envelope (5 Hz LPF version of the rectified raw EMG signal) used for visual biofeedback (Fig. 1). The raw EMG signal was also processed by the DSP to create two EMG envelopes (1 and 5 Hz LPF versions) that were used to modulate the EL fundamental frequency and EL activation/termination, respectively. Activation and termination thresholds were set independently, with the termination threshold set at 60–70% of the activation threshold level to facilitate uninterrupted EL voicing. The EL was mounted on a thick, flexible copper wire that bent once around the base of the neck and held the EL to the neck surface. Speech signals were recorded with a headset microphone (MPA III, AKG Acoustics, Vienna, Austria) located approximately 5 cm from the mouth. A photocell was mounted on the video monitor to synchronize visual stimulus presentation relative to recorded EMG and speech data. The raw EMG, audio, and photocell signals were recorded digitally (20 ks/s) with Axon Instruments hardware (Cyberamp 380, Digidata 1200) and software (Axon Instruments, Foster City, CA). Fig. 2 shows an individual using the EMG-EL, depicting the relative positions of the EL head and EMG electrodes.

2.3. Training protocol

The experiment occurred over 4 days for each participant. Naturally innervated strap muscle EMG was used to control the EMG-EL across two consecutive days, and RLN-innervated strap muscle EMG was used on the other two consecutive days, with the order of the muscle control source randomized. The 2 days focusing on one nerve control source were consecutive for all participants. One participant (S3) performed the experiment in four consecutive days, while the other two participants had 7 (S2) or 35 (S1) days between each 2-day block. The beginning of each testing day contained a setup period, during which the EMG electrodes were applied. An investigator then reviewed with the participant the basic device function, including the relationship between neck muscle contraction and EL output, and worked with the participant to obtain an appropriate activation threshold level. The first day of training for each strap muscle nerve supply (ANSA versus RLN) consisted of a series of Probes and unguided speaking sessions. Unguided sessions entailed participants practicing the speech tasks measured during the Probes and were conducted to see whether individuals could become proficient with the EMG-EL in a self-directed manner. The second day of training for each nerve supply similarly consisted of a series of Probes and practice sessions, but with the practice sessions guided by an investigator. These guided speaking sessions represented a more formal form of training for EMG-EL speech skill acquisition (see below). The detailed experimental protocol schedule is shown in Table 1. Six Probes were conducted each day to periodically sample participants' proficiency in EMG-EL speech as they practiced using the device in an unguided (day 1) or guided (day 2) manner. The Probes and speaking sessions included four speech tasks based on the work of Goldstein et al. (2007): vowels, sentences, phrases, and questions. Each Probe contained approximately ten tokens of vowels, six tokens of sentences, eight tokens of phrases, and ten tokens of question/statement pairs. The four speech tasks were presented in random order for each Probe.

Table 1
Schedule of experimental protocol.

Day 1: morning (control source 1)	Time (min)	Day 1: afternoon (control source 1)	Time (min)	Day 2: morning (control source 2)	Time (min)	Day 2: afternoon (control source 2)	Time (min)
Initial Instructions	5						
Hook up	10	Hook up	5	Hook up	5	Hook up	5
Set threshold	5	Set threshold	2	Set threshold	2	Set threshold	2
		Unguided Speaking	15			Guided Speaking	15
Probe 1	10	Probe 4	10	Probe 7	10	Probe 10	10
Unguided speaking	15	Unguided speaking	15	Guided speaking	15	Guided speaking	15
Probe 2	10	Probe 5	10	Probe 8	10	Probe 11	10
Break	5	Break	5	Break	5	Break	5
Unguided speaking	15	Unguided speaking	15	Guided speaking	15	Guided speaking	15
Probe 3	10	Probe 6	10	Probe 9	10	Probe 12	10
Lunch break	xx			Lunch break	xx		

The guided speaking sessions contained 15 min of focused practice on the tasks assessed during the Probes. Areas of focus included tasks that were perceived by the participant and investigators as more difficult for the participant. To ensure that each task was reviewed, every category was trained for at least one-half of a guided speaking session. Therefore, two sessions of guided speaking session time contained half-sessions of each of the four categories. The other three sessions were left to the choices of the participant and investigator. During the guided speaking sessions, the investigator directed the participant's attention to his or her speech quality as they spoke with the EMG-EL, as well as visual feedback of the EMG (Fig. 1).

A set of training strategies was established to use with the participants to improve performance. For example, when participants quickly or forcefully closed their mouths to signal the end of an intended vocalization, the EMG-EL often buzzed for several hundred milliseconds afterward, presumably due to generalized activation of neck musculature. However, if the participants gently closed their mouths, they were often able to terminate the EMG-EL voicing more quickly. Focusing on the visual biofeedback of the EMG envelope allowed participants to see the relationship between EMG level and EMG-EL activation/termination, and seemed to help participants terminate voicing more appropriately. During the phrases, participants were typically able to place appropriate breaks in their EMG-EL voicing by pausing after each line of the stimulus while attending to their EMG levels on the monitor.

To increase the dynamic range of pitch changes, different strategies were used for the different nerve control sources. When the RLN-innervated straps were the EMG-EL control source, the participants were instructed to imagine raising their pitch or increasing their vocal amplitude; however, when the naturally innervated straps were the control source, the participants were asked to imagine that they were lowering the pitch in order to raise the EMG-EL fundamental frequency. This difference in instruction was based on the observation that healthy normal strap muscle activation is consistently high during low-frequency vocalizations (see Vilkman, Sonninen, Hurme, & Korkko, 1996 for review) and that in the healthy larynx, RLN-innervated intrinsic laryngeal muscle (cricothyroid) activation is more frequently associated with pitch raising than lowering (e.g. Roubeau, Chevrie-Muller, & Lacau Saint Guily, 1997). The linear relationship between EMG envelope level and EMG-EL fundamental frequency was intuitive for participants, and their fundamental frequency control seemed to benefit from visual feedback of EMG envelope in real-time during EMG-EL speech in guided speaking sessions.

2.4. Stimulus materials

Different stimulus materials were used for the four speech tasks of vowel control, intelligibility, fluency, and intonation. The specific stimuli for intelligibility, fluency, and intonation that were presented during the guided and unguided sessions were different than those presented during the Probes to eliminate practice effects.

Vowel initiation, duration, and termination were measured to obtain objective data regarding participant ability to precisely control the timing of EMG-EL voicing. Instructions were given to the participant via visual commands with different background colors displayed on a video monitor. The presentation and scoring were based on the strategy reported by Goldstein et al. (2007). Instructions were presented to direct participants through the four phases of each token: rest, preparation, vocalization, and termination. The rest phase was 10 s, the stop phase 2 s, and the ready and

Table 2
Examples of fluency categories.

Category	Example phrase
Questions	Was there nothing to look at?
	No people too great?
	Did nothing excite you?
	Or make your heart beat?
Quotations	It poured and it lightninged.
	It thundered. It rumbled.
	"This isn't much fun."
	The poor elephant grumbled.
Carrier phrases	I do not like them with a fox.
	I do not like them in a box.
	I will not eat them in a house.
	I do not like them with a mouse.
Other	I have no time for tricks.
	I must go back and dig.
	I can't have you in here.
	Eating cake like a pig!

vocalization periods were randomly varied (1–2 and 2–4 s, respectively) to minimize anticipation of vowel sound start/ stop commands.

To estimate intelligibility, word pairs from the diagnostic rhyme test (DRT) (Voiers, 1977) were placed in a carrier sentence and read by the participants. The DRT contains 192 monosyllabic words presented in pairs that vary only in word-initial sounds. There are six categories within the DRT that differ by distinctive features: voicing, nasality, sustenation, sibilation, graveness, and compactness (see Voiers, 1983) for review of this metric). The corpus was split in half, allotting 48 pairs each for testing and training stimuli. Matched pairs from each of the six categories were presented to participants in the carrier phrase, "Write___again." to provide a consistent and more natural connected speech context.

Phrases were used to determine fluency of speech of the EMG-EL device. Phrases were selected from Dr. Seuss books for children (e.g. Cat in the Hat). Sentences from Dr. Seuss were chosen because they are at an appropriate reading level to facilitate fluency, and their rhyme and prosodic pattern induce a more melodic cadence when read than sentences typically used for such testing. The phrases were categorized by type: questions, quotations, carrier phrases, and other. Examples of categories are shown in Table 2. Each phrase was four to six lines long. Each of the four categories was represented twice in each Probe, making eight tokens per Probe. Repetitions were allowed for phrases if a non-linguistic error occurred, such as a cough, sneeze, or outside sound source.

Intonation was measured by comparing question/statement production of sentences. Three-word sentences were created containing monosyllabic words all beginning with voiced phonemes (e.g. Ben rang Barb, Grace dug bulbs, etc.) to avoid word-initial complications for the EL users. These sentences were constructed to mimic the syllable structure of the "Bev loves Bob" stimuli used by Gandour and Weinberg (1984). The presentation order of question and statement forms for each token was randomized, and participants were asked to repeat any sentence that they produced disfluently, since the focus was on pitch control rather than the other tested aspects of device control.

2.5. Data analysis

2.5.1. Vowel control

The scores for the vowel protocol included initiation, duration, and termination of production. Using custom software written in MATLAB we performed an automated analysis of EMG-EL vowel initiation, termination and duration using the microphone signal in relation to screen commands from the photocell (see Section 2.2). Vowel initiation was measured as the time difference between instruction to start, and initiation of vocalization. Termination was measured as the time difference between instruction to stop and the termination of vocalization. Duration was measured as the percent of time participants produced EMG-EL voice once they had initiated voicing after the screen

prompt for vowel production (''say/a/'') had appeared, and until the command to terminate voicing was given by the software (after a variable vowel production command interval of 2–4 s). The duration value was therefore reduced from 100% by any breaks in EMG-EL output after voice was initiated and by voice termination prior to the stop command. This method avoided duration scores being affected by reaction times, which were already represented in the vowel initiation score. Instances where EMG-EL voice was initiated prematurely (in the ''ready'' period before the screen prompt for vowel production appeared) were not included in the average duration measures.

2.5.2. Intelligibility

Intelligibility was measured by scoring DRT target words via listener comprehension. The sentences and targets were presented independently to three listeners. The listeners heard one sentence containing the target word played through headphones from a PC and had the target word and its matched pair on a printed sheet. Listeners were instructed to select which word from each word pair they heard spoken in each presented sentence, listening to the sentence as many times as needed. The average percentage of correct responses was compared across the six categories within the DRT. Ten percent of judged trials were repeated at least once. Intrajudge reliability was obtained by comparing the initial and repeated 10% of judged response, with an average exact agreement statistic for the three listeners of 94%. Interjudge reliability was calculated using the exact agreement statistic (87%) and the intraclass correlation statistic (ICC; 71%).

2.5.3. Fluency

The phrase fluency was scored by the first author. The number of sounds within each phrase were counted and annotated prior to listener presentation. Error types were labeled as the following: unfinished word, unfinished phrase, hesitation or block, prolongation, devoicing mid-word, and extended voicing. Fluency was calculated as a percentage of sounds performed accurately of the given total sound count. Sounds that were repeated, dropped, delayed in onset, or otherwise demonstrated a lack of control were counted as errors within the phrase. Dysfluencies unrelated to EMG-EL control, such as whole word repetitions (mostly attributable to reading difficulties), coughs, sneezes, or other outside factors were noted during the experiment, but were not incorporated in the scoring. Reliability of phrase fluency scoring was calculated by comparing fluency percentages of a subset of phrase trials (12.5%). The first author judged all trials and repeated the randomized subset. In addition, another Speech-Language Pathologist judged the same randomized subset of trials. The judgments of phrase fluency of the first author had an intrarater reliability score of 99% as measured with Pearson's R. The tokens scored by both the first author and the second listener had a Pearson's R of 82%.

2.5.4. Intonation

Statement and question intonation was judged by comparing the lowest EL fundamental frequency during the first 56% of the sentence to the highest fundamental frequency achieved during the last 44% of the sentence. Majewski and Blasdell (1969) found that listeners required pitch of a tone to change from 90 to 150 Hz for it to be considered a question significantly more often than chance. In preliminary experiments, we observed that a rise in EL voice from 90 to 150 Hz was sufficient for listeners to deem the utterance a question an average of 91.8% of the time (unpublished observations). Therefore, baseline pitch for the EMG-EL device was set at 90 Hz and needed to reach or exceed 150 Hz within the last 44% of the sentence to be scored as a question.

3. Results

3.1. Onset, duration, termination of vowels

Overall, differences in the participants' control capabilities for vowel initiation, duration, and termination using either nerve supply as a control source varied as a function of participant and task. Fig. 3 shows the vowel initiation, duration, and termination performance of participants over training Probes 1–6 (unguided testing/training day) and Probes 7–12 (guided testing/training day). Individual t-tests were performed between pre-training performances (Probe 1) of the two nerve supplies, with the family-wise alpha of 0.05 with the Bonferroni adjustment due to multiple t-tests (24 per participant) leading to a comparison-wise alpha of 0.002 for significance.
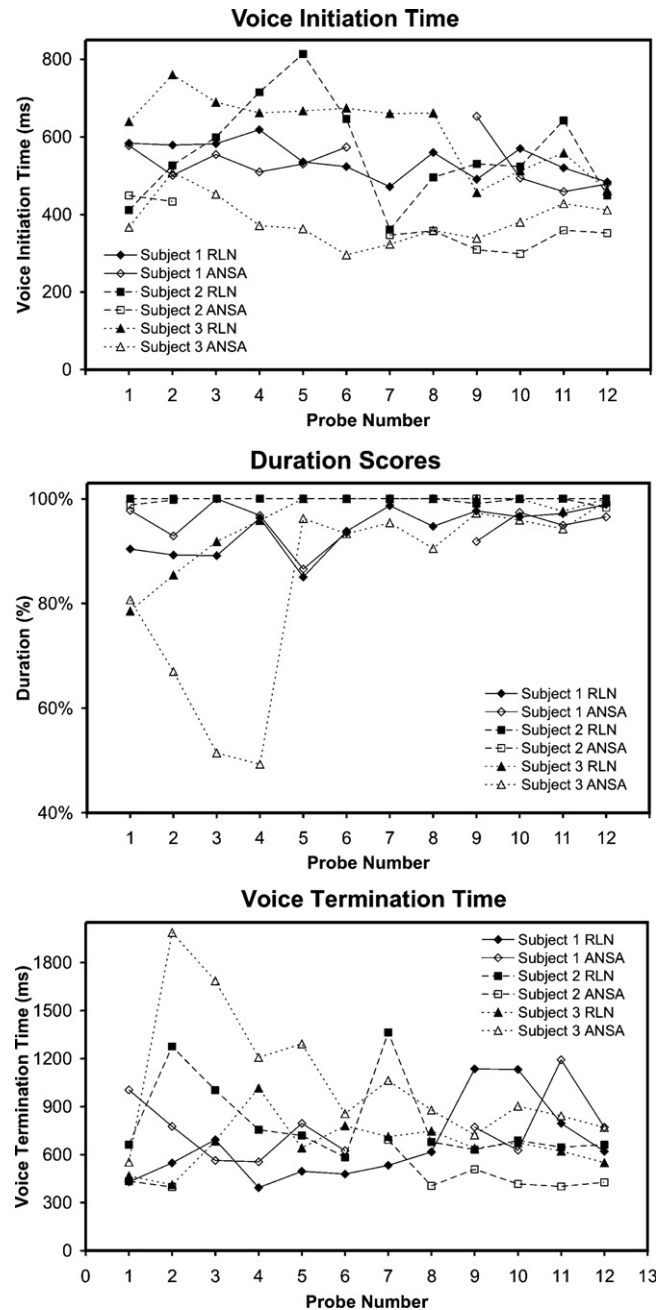
Fig. 3. Voice onset time, duration scores, and termination time are plotted for both the RLN-innervated and ANSA-innervated EMG-EL control locations across Probes 1–12 for each participant. No consistent difference was found between control locations or between the unguided day of testing (Probes 1–6) versus the guided day of testing (Probes 7–12) on all measures (see also Figs. 3–5).

Because each probe consisted of ten vowel prompts, in general, *t*-tests were performed with $N = 10$ (i.e. treating trials as independent observations). However, because successful voice initiation and duration must be accomplished in order to estimate the VTT, the number of values used for *t*-tests varied based on participant performance. No significant differences based on nerve supply (two-tailed) were found in pre-training VIT, duration, or VTT, with the exception of participant S3, whose VIT pre-training was significantly lower (better) when controlling the device with his ANSA-innervated side ($p = 0.001$). However, participant S3 had poor pre-training VIT and duration performance

such that there were not enough samples of VTT to perform a *t*-test. Post-training (Probe 12) differences in VIT, duration, and VTT between nerve supplies were also non-significant, with the exception of participant S2, whose VTT post-training was significantly lower (better) when controlling the device with his ANSA-innervated side ($p = 0.0001$).

The training study using the EMG-EL in individuals utilizing RLN-innervated strap muscle by Goldstein et al. (2007) defined ''success'' of vowel initiation and termination based on the percentage of tokens produced within a predetermined criterion (VIT criterion of 390 ms, VTT criterion of 330 ms). Their participants ($N = 3$) had initial vowel initiation scores ranging from 0 to 10%, and post-training scores ranging from 10 to 100%. When scored in this manner, the participants in this study had initial vowel initiation scores ranging from 0 to 40% when using their RLN-innervated strap muscles, and 0–56% when using their naturally innervated strap muscles. After training, these individuals had vowel initiation scores ranging from 10 to 50% using their RLN-innervated strap muscles, and 30 to 70% using their naturally innervated strap muscles. The participants studied by Goldstein et al. all had initial and post-training vowel termination scores of 0%. The participants in this study had initial vowel termination scores ranging from 0 to 14% when using their RLN-innervated strap muscles, and 0 to 11% when using their naturally innervated strap muscles. After training, these individuals had vowel termination scores ranging from 0 to 11% using their RLN-innervated strap muscles, and 0 to 10% using their naturally innervated strap muscles.

To assess possible learning, each participant's performance in Probe 1 was compared to Probe 12 using paired *t*-tests (one-tailed). No significant difference was found between pre- and post-training performance in voice initiation, duration, or termination for any participant or nerve supply. The associated *p*-values are shown in Table 3. Again, participant S3 had poor pre-training VIT and duration performance such that there were not enough samples of pre-training VTT to perform a *t*-test between pre- and post-training.

3.2. Intelligibility

The three listening judges were able to choose the correct word each EMG-EL speaker produced with accuracy consistently above chance for all categories. The categories of sustenation, voicing and graveness had the lowest listener accuracy, with overall averages across participants of 82, 69 and 84%, respectively. The other four DRT categories had average listener accuracies greater than 94%. Overall, listener accuracy was an average of 86% across the three participants. Individual *t*-tests were performed between pre-training performances (Probe 1) of the two nerve supplies (two-tailed, d.f. ≥ 22), post-training performances (Probe 12) of the two nerve supplies (two-tailed, d.f. ≥ 22), as well as between pre- and post-training performance within each nerve supply (one-tailed, d.f. ≥ 22). The intelligibility ratings were consistently high and did not vary significantly with training or nerve supply (see Fig. 4). No significant differences were found for any participant, with all *p*-values found to be larger than 0.002.

3.3. Fluency

Per listener ratings, the participants' speech was consistently fluent; the range of fluencies found for any individual regardless of control source or probe number was 90.9–99.6%. Hesitation, part-word devoicing, and whole word devoicing were the most frequent fluency errors across the participants. Hesitation was defined as an extended pause either between words or within words. Part-word devoicing was labeled when the EMG-EL device inappropriately stopped buzzing for a portion of a word. Whole word devoicing occurred when an entire word was not voiced. Typically, attempted articulation was audible regardless of whether the EMG-EL produced a voice, allowing listeners to identify this fluency error.

Table 3
Pre- and post-training vowel comparison *p*-values.

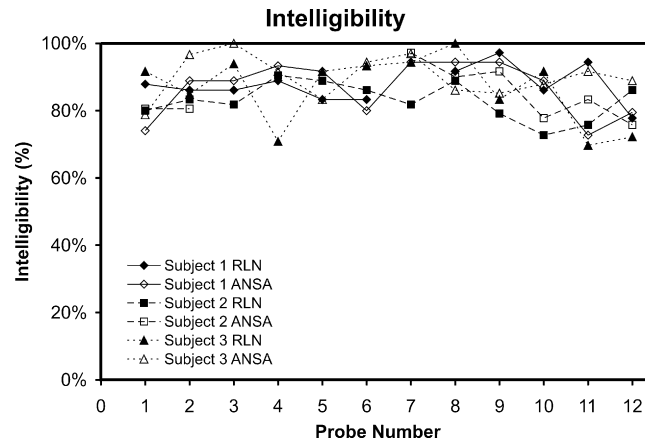| Vowel production	RLN control			ANSA control		
	S1	S2	S3	S1	S2	S3
Initiation	0.038	0.275	0.008	0.095	0.045	0.214
Duration	0.006	0.500	0.005	0.006	0.405	0.052
Termination	0.500	0.500	0.191	0.087	0.477	N/A

Fig. 4. Speech intelligibility is plotted for both the RLN-innervated and ANSA-innervated EMG-EL control locations across Probes 1–12 for each participant.
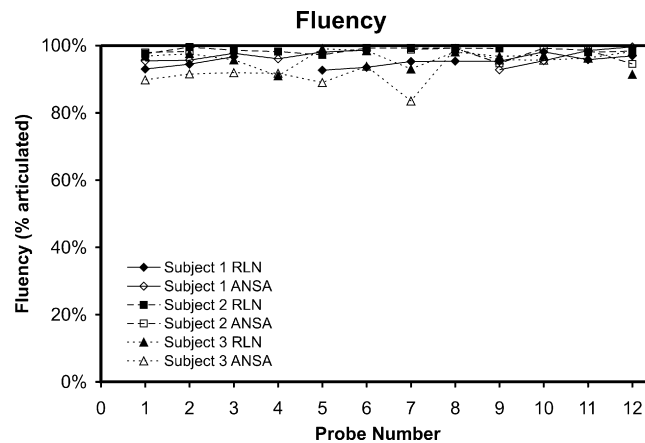

Fig. 5. Speech fluency is plotted for both the RLN-innervated and ANSA-innervated EMG-EL control locations across Probes 1–12 for each participant.

When individual t-tests were performed between pre-training performances (Probe 1) of the two nerve supplies (two-tailed, d.f. = 14), post-training performances (Probe 12) of the two nerve supplies (two-tailed, d.f. = 14), as well as between pre- and post-training performance within each nerve supply (one-tailed, d.f. = 14), the fluency ratings did vary significantly with training or nerve supply in a few cases. Fig. 5 shows fluency performance averages as training progressed. Participant S1 and S3 had a significant improvement in the ANSA-innervated control of fluency between pre- and post-training ($p = 0.002$, $p = 0.0001$, respectively). Participant S3 also showed a difference between pre-training nerve-supplies ($p = 0.002$), with the RLN-innervated side showing greater control.

3.4. Intonational contrasts

The participants produced fluctuations in pitch throughout their sentences, yet had difficulty consistently differentiating questions versus statements through intonational contrasts (see Methods). Fig. 6 shows the percent of successful question/statement intonations. As a group, the participants produced intonation appropriate for attempted questions and statements on an average of 75% of trials by the final Probe. However, individual t-tests performed between pre- and post-training performance within each nerve supply (one-tailed, d.f. $\geq$ 32) failed to find significant differences for any participant, with all p-values found to be larger than 0.002. The same was true for individual t-tests between pre-training performances (Probe 1) of the two nerve supplies (two-tailed, d.f. $\geq$ 32) and post-training performances (Probe 12) of the two nerve supplies (two-tailed, d.f. $\geq$ 32, with all p-values found to be larger than
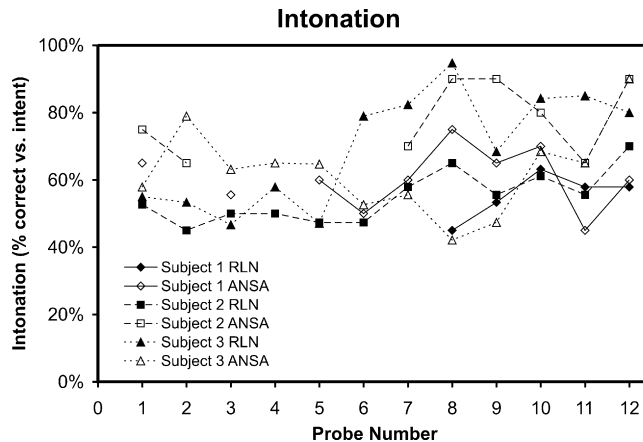
Fig. 6. Speech intonation is plotted for both the RLN-innervated and ANSA-innervated EMG-EL control locations across Probes 1–12 for each participant. Although no participants showed pre- to post-training learning with statistical significance less than the Bonferroni adjusted alpha level, participant S3 did show a trend of increased intonation between pre- and post-training for both ANSA and RLN sides (one-tailed t-tests; ANSA, $p = 0.012$, RLN, $p = 0.048$).

0.002. The pre-training intonation data of participant S1 was partially corrupted by an inappropriate equipment setting, which precluded statistical analysis of his pre-training intonation compared between nerve supplies, as well as analysis of his pre- versus post-training RLN intonation control.

4. Discussion

Three individuals undergoing total laryngectomy were prospectively studied for their ability to control an EL using neck surface EMG signals from RLN-innervated and naturally innervated strap muscles. All three participants were able to speak relatively intelligibly and fluently with the EMG-EL after receiving only basic instruction before their very first Probe. Performance did not consistently differ for the two control source nerve supplies, nor did performance consistently improve across the 2 days of testing for either control source. Specifically, vowel onset, duration, termination, and speech intelligibility, fluency, and pitch modulation capabilities using the EMG-EL did not systematically differ between the RLN and natural strap muscle nerve supplies before or after training, and none of these dependent variables significantly improved from the first testing Probe (1) versus the last Probe (12; at the end of the second day) for either nerve supply, with the exception of speech fluency using the ansa-innervated strap muscle recording location for two of the three participants.

4.1. EMG-EL training effects

Our prior study of training effects on speech production using the EMG-EL (Goldstein et al., 2007) demonstrated performance improvements attributed to training for anatomically intact participants using naturally innervated neck strap muscles for EMG-EL control, as well as for four individuals using RLN-innervated strap muscles after total laryngectomy. Our present participants spoke remarkably well with the EMG-EL in their initial test Probe, immediately producing relatively intelligible, fluent speech and thereby resembling the post-training speech of our prior cohort in terms of their ability to successfully read aloud words, sentences, and a paragraph (speech intelligibility and fluency were not directly measured in the prior study). Listeners in the present study had some difficulty distinguishing minimal difference word pairs for EMG-EL speech samples in the categories of sustenation (distinguishing stops versus fricatives) and voicing (distinguishing voiced versus voiceless sounds), which are known challenges in EL speech due to the loss of DC air flow and rapid voice onset/offset control, respectively. Nevertheless, our participants were able to speak in the high end of the reported EL speech intelligibility range (32–90%; for review see Hillman, Walsh, & Heaton, 2005) from the very beginning of their testing/training (see Fig. 4), and fluency was likewise high from the very first use of the EMG-EL device. These high initial scores likely explain why performance improvements were not obtained with training for our small group of participants.

The discrepancy in initial EMG-EL performance between the present participants and our prior cohort likely stems from differences in how the device activation/termination thresholds were set and what initial instructions the participants received. The prior participants were not told the relationship between neck muscle contraction and device activation (e.g. stronger contraction activates device more often and produces higher vocal pitch) before their initial Probes, and the EMG-EL activation threshold was uniformly set in relation to the maximum voice-induced contraction (RMS EMG envelope) at a level that may have been too high for consistent device activation. For the present study, participants were explicitly shown the relationship between neck muscle activation (e.g. vocal effort) and device function prior to the first Probe, and thresholds were individualized through iterative adjustment to produce the best speech result prior to each Probe. Moreover, the original analog version of the EMG-EL device provided a termination threshold based on an internal (fixed) activation-threshold-dependent hysteresis band, whereas the present digital version of the EMG-EL provided control of the termination threshold as an adjustable percentage of the activation threshold. Therefore, as with the activation threshold in the present report, the termination threshold was individualized through iterative adjustment to better facilitate sustained voicing while minimizing unintentional voice prolongation for each participant. A combination of these factors could account for the stronger initial speech capabilities for the present participants compared to our prior cohort.

Although participants tended to have a high initial level of EMG-EL speech proficiency which they then maintained across successive probes, significant improvements were observed in speech fluency for the natural nerve supply control source in two of the three participants (S1 and S3). At the time of the final Probe, however, the two control sources did not significantly differ for any of the measured parameters, indicating equivalent control capabilities for the natural and RLN nerve supplies after training. Moreover, for both nerve supplies there was a general trend for improved intonation control with training. Participants' ability to successfully intone an interrogative versus statement appeared relatively constant across Probes 1–6 on the first (unguided) day of device use, improving in performance after guided (trained) use of the EMG-EL commenced on the second day (from Probe 7 onward; see Fig. 6). Therefore, the ability to intone questions versus statements may have benefitted from formal training for both RLN and natural strap muscle nerve supplies. This was not unexpected, considering that vocal intonation was one of the most difficult skills to acquire in our earlier study of EMG-EL training effects (Goldstein et al., 2007) for both the natural and RLN strap muscle nerve supplies, and showed the most direct correspondence between improvements in performance and the initiation of training. The present participants may have realized further improvements in performance with additional training beyond their single intensive day, as some of our prior participants had conspicuous leaps in intonation control towards the end of the training protocol, which was spread out in shorter time increments across several days.

4.2. Targeted muscle reinnervation for prosthesis control

An ideal EMG control source for any prosthesis would be one that naturally relates to the previous functions of the lost anatomy, providing a physiologically relevant and therefore highly intuitive control mechanism. Recent efforts have been made to obtain such an optimal EMG control source for an advanced arm prosthesis by transferring residual nerves of the amputated arm to host muscles in the adjacent chest region (Kuiken, 2006; Kuiken et al., 2007; Zhou et al., 2007). This approach, known as TMR, has been successfully performed in four individuals to date (Kuiken et al., 2007), substantially improving their prosthetic limb function and providing an intuitive control mechanism which requires much less effort and attention than prior EMG control options.

The RLN transfer performed at the time of laryngectomy in our participants was a form of TMR, with neck strap muscles that had been rendered mechanically nonfunctional at the time of laryngectomy being used as a biological amplifier of RLN motor commands. In this study, comparison of prosthetic voice control capabilities between the RLN-innervated strap muscles versus naturally innervated strap muscles tested the hypothesis that the RLN would provide a better control source because it would presumably convey more precise vocal-related activity than what is normally found in neck strap muscles. Our findings do not support this hypothesis, as TMR of the RLN did not provide an advantage over the natural strap muscle nerve supply for control of the EMG-EL. There are several possible reasons why this anticipated advantage of RLN control was not observed, primarily including our EMG-EL signal processing strategy and the natural presence or propensity for vocal-related activity in neck strap musculature as discussed below.

The EMG-EL prosthesis used by participants in this study had a simple dual-thresholding strategy of the EMG root-mean-squared (RMS) amplitude envelope to determine voice initiation and termination, combined with a proportional

relationship between suprathreshold EMG and vocal fundamental frequency. This combination of EMG envelope amplitude thresholding and proportional control is common for EMG-controlled limb prosthesis function (for review see Kuiken, 2006), but does not take advantage of the rich information content provided by TMR when neural drive of diverse muscle groups are expressed in a single host muscle or single recording location. For example, Zhou et al. (2007) demonstrated recognition of 16 intended arm, hand, and finger/thumb movements after post-processing surface EMG signals from individuals having undergone TMR after arm amputation. Although their limb control capabilities after TMR based on a simple real-time amplitude-based EMG analysis was clearly superior to conventional EMG control, their control could have been much more dexterous if the neural drive to the dozens of hand and arm muscle were extracted more independently from the surface EMG signals rather than lumped together in the signal amplitude analysis.

Similarly, the RLN normally supplies a diverse set of laryngeal muscles including muscles that tense or adduct the vocal folds (thyroarytenoid, lateral cricoarytenoid, and interarytenoid) and the single muscle which acts antagonistically to these others and abducts the vocal folds (posterior cricoarytenoid). Amplitude-based EMG analysis of the combined motor drive to these agonist–antagonist muscle groups is probably not the best way of extracting precise vocal-related activity after RLN TMR. For example, brief voiced/voiceless contrasts or rapid voice termination in general may be better obtained through more sophisticated signal processing strategies than envelope thresholding, just as numerous hand/arm functions are discernable using pattern classification techniques after TMR that are not evident in the EMG envelopes alone (Zhou et al., 2007). Therefore, although strap muscle activation with a natural nerve supply supports amplitude-based prosthetic voice production comparable to RLN innervation, the RLN TMR procedure might provide vocal-related information not currently utilized by our EMG-EL prosthesis. Future post-processing experiments with our current dataset may reveal additional vocal-related information from the RLN-innervated strap muscles, offering justification for the extra surgical procedures required to perform RLN TMR during total laryngectomy. However, we cannot recommend TMR of the RLN for amplitude-based EMG control of an electrolarynx in light of the present findings.

The main advantage of TMR is that it provides an intuitive neural control source that physiologically relates to the lost function being prosthetically restored. TMR of the RLN in the present study might not have produced a distinct advantage over the natural nerve supply for preserved neck strap muscles due to the inherent overlap of laryngeal (via RLN) and strap muscle activation during phonation. The group of strap muscles preserved in our three participants included the sternohyoid, sternothyroid and omohyoid, which are known to contract during phonation—particularly when producing loud or low-frequency voice (see Vilkman et al., 1996 for review). Given that our participants were able to speak proficiently using naturally innervated strap muscles prior to formal EMG-EL training or visual EMG biofeedback, it demonstrates inherent and/or easily acquired vocal-related activity patterns from these muscles.

Maintaining the natural nerve supply to preserved strap muscles instead of performing RLN TMR avoids the risk of a failed neurorrhaphy with the RLN, prevents denervation atrophy during the post-surgical reinnervation period, and provides an immediately available EMG-EL control source instead of needing to wait 6 or more months for the RLN TMR to become effective (Heaton et al., 2004). Moreover, strap muscle preservation adds little additional time to the typical laryngectomy surgery and does not sacrifice the function of any muscles, so it would be a reasonable option (when oncologically appropriate) for providing an EMG-EL control source. However, other neck and tongue-base musculature remaining after total laryngectomy may likewise provide a useful EMG-EL control source, obviating the need to intentionally preserve or re-position any strap musculature during surgery. We are currently investigating alternative EMG-EL control sources in neck surface recordings of individuals who have undergone standard total laryngectomy surgery to explore this possibility.

5. Conclusion

In a small set of participants we have shown that preserved neck strap muscles with either their natural nerve supply or a transferred RLN nerve supply can serve as an effective control source for a hands-free EMG-controlled electrolarynx (EMG-EL). High initial device proficiency likely precluded consistent improvements in EMG-EL control after training for all measured parameters. TMR of the RLN to neck strap muscles did not provide an advantage over preservation of the natural nerve supply, suggesting that RLN transfer is unnecessary for effective EMG-EL control. Alternative neck and face recording locations are being explored to see if the EMG-EL can be utilized by individuals who have undergone conventional laryngectomy surgery without special nerve or muscle preservation.

Appendix A. Continuing education

1. How is the pitch of the EMG-EL controlled?
   A. The pitch is correlated to the frequency content of the EMG, with higher frequency EMG corresponding to higher fundamental frequency (F0).
   B. The pitch is correlated to the level of suprathreshold EMG energy, with greater EMG energy corresponding to higher fundamental frequency (F0).
   C. The pitch is correlated to the level of suprathreshold EMG energy, with greater EMG energy corresponding to lower fundamental frequency (F0).
   D. The pitch is manually controlled by the user of the EMG-EL with a button switch.
2. The Diagnostic Rhyme Test (DRT) was used as an estimate of:
   A. Intelligibility.
   B. Fluency.
   C. Prosody.
   D. Naturalness.
3. Which of the following is a neck strap muscle?
   A. Cricothyroid muscle.
   B. Sternocleidomastoid muscle.
   C. Trapezius muscle.
   D. Sternothyroid muscle.
4. In the healthy larynx, activation of the cricothyroid muscle:
   A. Closes the glottis.
   B. Opens the glottis.
   C. Is associated with pitch raising.
   D. Is associated with pitch lowering.
5. Here, the acronym TMR stands for:
   A. Targeted muscle reinnervation.
   B. Timed muscle reactivation.
   C. Triggered myopathy research.
   D. Trained modification research.

References

Cancer facts & figures. (2008). Atlanta: American Cancer Society.

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). Austin, TX: PRO-ED.

Gandour, J., & Weinberg, B. (1984). Production of intonation and contrastive stress in electrolaryngeal speech. *Journal of Speech and Hearing Research, 27*(4), 605–612.

Goldstein, E. A., Heaton, J. T., Kobler, J. B., Stanley, G. B., & Hillman, R. E. (2004). Design and implementation of a hands-free electrolarynx device controlled by neck strap muscle electromyographic activity. *IEEE Transactions on Biomedical Engineering, 51*(2), 325–332.

Goldstein, E. A., Heaton, J. T., Stepp, C. E., & Hillman, R. E. (2007). Training effects on speech production using a hands-free electromyographically controlled electrolarynx. *Journal of Speech, Language, and Hearing Research, 50*(2), 335–351.

Gray, S., & Konrad, H. R. (1976). Laryngectomy: Postsurgical rehabilitation of communication. *Archives of Physical Medicine and Rehabilitation, 57*(3), 140–142.

Heaton, J. T., Goldstein, E. A., Kobler, J. B., Zeitels, S. M., Randolph, G. W., Walsh, M. J., et al. (2004). Surface electromyographic activity in total laryngectomy patients following laryngeal nerve transfer to neck strap muscles. *Annals of Otology, Rhinology, & Laryngology, 113*(9), 754–764.

Hillman, R. E., Walsh, M., & Heaton, J. T. (2005). Laryngectomy speech rehabilitation: A review of outcomes. In P. Doyle & R. L. Keith (Eds.), *Contemporary consideration in the treatment and rehabilitation of head and neck cancer. Voice, speech, and swallowing* (pp. 75–90). Austin: PRO-ED.

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 Otology, Rhinology, & Laryngology Supplement, 172*, 1–27.

Kuiken, T. (2006). Targeted reinnervation for improved prosthetic function. *Physical Medicine and Rehabilitation Clinics of North America, 17*(1), 1–13.

Kuiken, T. A., Miller, L. A., Lipschutz, R. D., Lock, B. A., Stubblefield, K., Marasco, P. D., et al. (2007). Targeted reinnervation for enhanced prosthetic arm function in a woman with a proximal amputation: A case study. *Lancet, 369*(9559), 371–380.

Majewski, W., & Blasdell, R. (1969). Influence of fundamental frequency cues on the perception of some synthetic intonation contours. *The Journal of the Acoustical Society of America, 45*(2), 450–457.

Meltzner, G. S., Hillman, R. E., Heaton, J. T., Houston, K. M., Kobler, J. B., & Qi, Y. (2005). Electrolaryngeal speech: The state of the art and future directions for development. 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. 571–590). Austin, TX: PRO-ED.

Mendenhall, W. M., Morris, C. G., Stringer, S. P., Amdur, R. J., Hinerman, R. W., Villaret, D. B., et al. (2002). Voice rehabilitation after total laryngectomy and postoperative radiation therapy. *Journal of Clinical Oncology, 20*(10), 2500–2505.

Monahan, G. (2005). Clinical troubleshooting with tracheoesophageal puncture voice prostheses. 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. 481–502). Austin, TX: PRO-ED.

Morris, H. L., Smith, A. E., Van Demark, D. R., & Maves, M. D. (1992). Communication status following laryngectomy: The iowa experience 1984–1987. *The Annals of Otology, Rhinology, and Laryngology, 101*(6), 503–510.

Roubeau, B., Chevrie-Muller, C., & Lacau Saint Guily, J. (1997). Electromyographic activity of strap and cricothyroid muscles in pitch change. *Acta Otolaryngologica, 117*(3), 459–464.

Vilkman, E., Sonninen, A., Hurme, P., & Korkko, P. (1996). External laryngeal frame function in voice production revisited: A review. *Journal of Voice, 10*(1), 78–92.

Voiers, W. D. (1977). Diagnostic evaluation of speech intelligibility. In M. E. Hawley (Ed.), *Speech intelligibility and speaker recognition*. Stroudsburg, PA: Dowden, Hutchinson, and Ross.

Voiers, W. D. (1983). Evaluating processed speech using the diagnostic rhyme test. *Speech Technology, 1*, 338–352.

Zhou, P., Lowery, M. M., Englehart, K. B., Huang, H., Li, G., Hargrove, L., et al. (2007). Decoding a new neural machine interface for control of artificial limbs. *Journal of Neurophysiology, 98*(5), 2974–2982.