Design and Implementation of a Hands-Free Electrolarynx Device Controlled by Neck Strap Muscle Electromyographic Activity

Ehab A. Goldstein*, James T. Heaton, James B. Kobler, Garrett B. Stanley, *Associate Member, IEEE*, and Robert E. Hillman

***Abstract*—The electrolarynx (EL) voice prosthesis is widely used, but suffers from the inconvenience of requiring manual control. Therefore, a hands-free EL triggered by neck muscle electromyographic (EMG) activity was developed (EMG-EL). Signal processing circuitry in a belt-mounted control unit transforms EMG activity into control signals for initiation and termination of voicing. These control signals are then fed to an EL held against the neck by an inconspicuous brace. Performance of the EMG-EL was evaluated by comparison to normal voice, manual EL voice, and tracheo-esophageal (TE) voice in a series of reaction time experiments in seven normal subjects and one laryngectomy patient. The normal subjects produced voice initiation with the EMG-EL that was as fast as both normal voice and the manual EL. The laryngectomy subject produced voice initiation that was slower than with the manual EL, but faster than with TE voice. Voice termination with the EMG-EL was slower than normal voice for the normal subjects, but not significantly different than with the manual EL. The laryngectomy subject produced voice termination with the EMG-EL that was slower than with TE or manual EL. The EMG-EL threshold was set at 10% of the range of vocal-related EMG activity above baseline. Simulations of EMG-EL behavior showed that the 10% threshold was not significantly different from the optimum threshold produced through the process of error minimization. The EMG-EL voice reaction time appears to be adequate for use in a day-to-day conversation.**

***Index Terms*—Electromyography, prosthesis, voice.**

I. Introduction

HUMAN voice is the foundation of self-expression and vocal communication with others. Unfortunately, each year thousands of people undergo a laryngectomy, which is the surgical removal of the larynx for the treatment of cancer or trauma. Loss of the larynx results in the inability to produce normal speech and the need to breathe through a hole (stoma) in the neck. However, since the main articulators are still intact, a prosthetic device can be used to acoustically excite the vocal tract for production of alaryngeal speech. The three main methods of alaryngeal speech are esophageal, tracheo-esophageal (TE), and electro-laryngeal (EL).

During esophageal speech, the patient swallows air to inflate the esophagus and then expels it to make the pharyngo-esophageal tissue vibrate, thus producing a belch-like voice. In TE speech, a one-way valve diverts air from the trachea into the esophagus to drive the pharyngo-esophageal tissue for phonation. The EL is a hand-held battery-powered transducer that injects a buzzing sound through the neck tissue or the oral cavity into the vocal tract. All three methods produce acoustic energy that replaces voice and excites the vocal tract to generate speech.

Esophageal speech has reduced loudness and requires lengthy training with only a small percentage of patients being capable of using it for verbal communication [1]. The use of the TE valve allows for more functional speech that is easier for patients to produce. However, due to a variety of surgical, morphological, and behavioral factors, only about a third of laryngectomy patients can use it successfully [1]. Thus, EL speech continues to play a major role in laryngectomy rehabilitation, with multiple studies reporting that over half of the laryngectomy patients use an EL for verbal communication [1]–[3].

A major inconvenience in using a conventional TE prosthesis or a hand-held EL device is that they occupy the use of one hand. Before being able to speak or respond to a question, a TE user usually has to manually occlude the stoma in the neck. Similarly, an EL user has to reach into his pocket for the device and apply it to the neck surface or oral cavity. A recent survey of EL users found that the inconvenience of use was ranked most problematic, followed by the monotonic nature of the speech [4]. These results indicated the potential benefit in providing hands-free EL control.

A hand-held EL is composed of three parts: a transducer that mechanically produces the sound, an electric circuit that drives the transducer, and a battery pack that provides the power. On/off control has been typically achieved using a button-switch. The system described here uses electromyographic (EMG) signals, a proven method in the fields of limb prostheses [5]–[9], and human-computer interface technology [10]–[12]. Electrical signals from neck muscles are detected using a surface electrode, and are processed to produce on/off control. This report covers device design and assessment of system performance relative to normal, EL, and TE alaryngeal voice. Furthermore, the optimum on/off threshold was investigated through simulation of device behavior and comparison with normal voice.

Manuscript received January 10, 2003; revised June 1, 2003. This work was supported in part by the W. M. Keck Foundation and in part by the Department of Veterans Affairs Rehabilitation Research and Development under Grant C2243-2DC. *Asterisk indicates corresponding author.*

*E. A. Goldstein is with the Harvard-MIT Division of Health Science and Technology, and the Division of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02143 USA (e-mail: ehab@post.harvard.edu).

J. T. Heaton, J. B. Kobler, and R. E. Hillman are with the Voice and Speech Laboratory at the Massachusetts Eye and Ear Infirmary, and the Department of Otology and Laryngology, Harvard Medical School, Boston, MA 02114 USA.

G. B. Stanley is with the Division of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138 USA.

Digital Object Identifier 10.1109/TBME.2003.820373

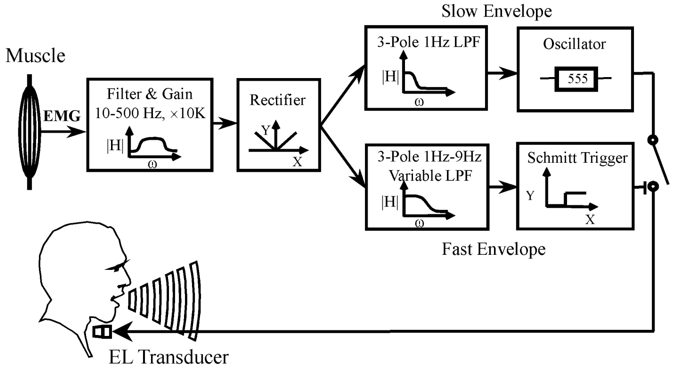

Fig. 1. Processing stages of the EMG-EL processor: EMG signal is filtered and amplified, then rectified and smoothed to produce a slow envelope (top path) to modulate pitch, and a fast envelope (bottom path) to control on/off of the EL.

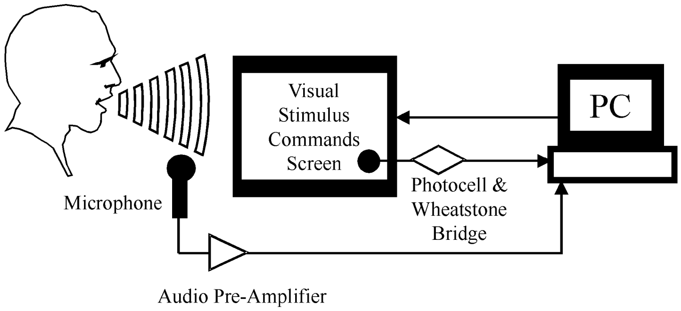

Fig. 2. Setup for measuring voice initiation and termination times. Visual stimulus was administered by the computer and displayed on a video monitor. Audio signal was recorded with a microphone, and stimulus signal was captured with a photocell circuit.

II. Methods

A. The EMG-EL System

A schematic representation of the EMG-controlled EL system (EMG-EL) is shown in Fig. 1. The EMG processing circuit produces an envelope waveform proportional to the time-averaged power in the EMG signal. The time-varying envelope controls on/off triggering and fundamental frequency (pitch) of a commercial EL transducer (NuVois, Mountain Precision Mfg., ID), which is held against the neck with a simple brace.

The electrical activity of the infra-hyoid neck strap muscles is detected on the surface of the skin using a bipolar electrode (DelSys Inc., Boston MA). The EMG signal amplitude is on the order of tens of microvolts, with most energy in the frequency range of 10 to 500 Hz. Therefore, the EMG signal is amplified and bandpass filtered to increase signal-to-noise ratio (SNR). The signal is then actively rectified and sent through two parallel pathways, each with a 3-pole low-pass filter (LPF). The first LPF has a corner frequency of approximately 1 Hz, while the second has a variable corner frequency that could be adjusted from 1 to 9 Hz. The two pathways produce two envelopes with "slow" and "fast" time constants.

The fast envelope is fed to a Schmitt trigger to turn the transducer on and off based on a controllable threshold voltage. Simultaneously, the slow envelope is used to control the fundamental frequency of EL voice by directly modifying the frequency of the oscillator driving the EL transducer. The other parameters of the EMG-EL processor are EMG gain, EL volume, starting fundamental frequency, and corner frequency for the fast-varying envelope. This paper will discuss voice initiation and termination, but not control of fundamental frequency.

B. EMG-EL Performance Assessment

To compare user performance with the EMG-EL to other voice sources, a reaction time protocol was adopted [13]–[16]. A visual cue signaled the subject to start and stop vocalizing, and the acoustic output was recorded. The time delay between the presentation of the visual cue and voice initiation and termination was used as a measure of vocal reaction time. Voice initiation time (VIT) and voice termination time (VTT) were measured for the EMG-EL, a conventional push-button EL, and normal voice in seven normal subjects, as well as TE voice in one laryngectomy subject.

The normal subjects were 4 males and 3 females, whose ages ranged from 21 to 49 years, averaging 27.5 years. The laryngectomy patient was a 61-year-old male who underwent a total laryngectomy 30 months prior to testing. The surgical procedure for this patient included preservation of the omohyoid strap muscles. The right and left omohyoid muscles were detached from the hyoid bone and were abutted and sutured together at the midline above the tracheostoma. The right omohyoid retained its original innervation, while the left omohyoid was selected for reinnervation via anastomosis of its motor nerve (left ansa cervicalis) to the left recurrent laryngeal nerve. Details of the surgical procedure and additional results will be presented in a separate publication. The laryngectomy subject initially used an EL, until he was fitted with a TE valve a few months after the surgery. The subject primarily used TE speech, but also used the EL on occasion whenever his TE valve was not functional.

Each subject was seated in a sound-treated chamber, with a video monitor (Sony Trinitron GVM 2020) placed one meter away. Speech signals were recorded with a condenser microphone (Sony ECM-50PSW) 15 cm from the mouth. The stimulus presentation, data acquisition and analysis were performed using a PC. Matlab (Mathworks, Natick, MA) software controlled the sound card (Aureal Semiconductor, Vortex AU8830 PCI) as a 2-channel input device. For timing accuracy, a photocell was mounted on the video monitor to detect the instant of visual stimulus presentation (see Fig. 2). The photocell output was recorded along with speech signals. Using the acoustic signal from the microphone and the electric signal from the photocell, the time delays between the stimulus and voice initiation and termination were measured with a resolution of 1 ms.

EMG signals were detected with a noninvasive DelSys DE2.1 surface electrode placed over the neck strap muscles. The EMG electrode was positioned along the midline of the neck, directly below the thyroid prominence of the normal subjects, and over the left omohyoid muscle above the stoma of the laryngectomy subject. An Ag/AgCl gel ground electrode (Kendall LTP, Chicopee, MA) was placed on the shoulder. The EMG electrode was connected to the EMG-EL to control initiation and termination of prosthetic voice.

Normal subjects received brief training on how to use the EL (Servox INTON, Cologne, Germany) and the EMG-EL, while the laryngectomy subject underwent a more extensive training protocol. The EMG gain was set so that the peak EMG envelope

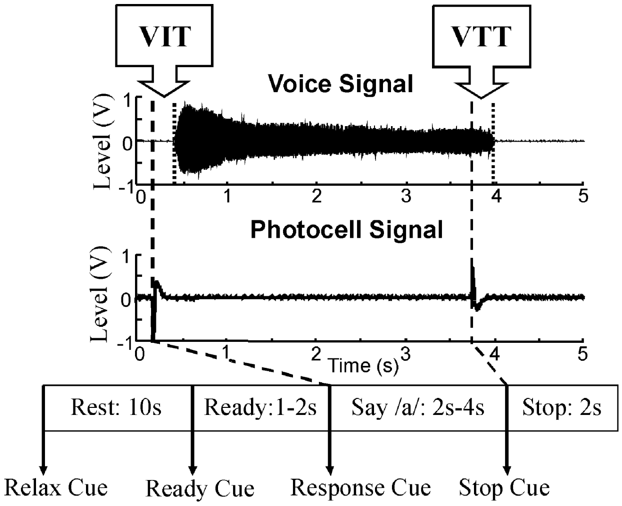

Fig. 3. Visual stimulus for the voice initiation and termination time experiment. Top trace shows audio signal, while bottom trace shows photocell signal indicating visual stimulus timing.

elicited by a low pitch vocalization was on the order of 500 mV. The electrode contact with the skin was adjusted until the resting baseline was less than 50 mV. Conductive electrode gel was used when necessary.

The EMG-EL threshold determined whether the sound transducer was on or off, based on the amplitude of the fast EMG envelope. Because different subjects had varying levels of baseline and vocal-related neck strap muscle EMG activity, the threshold was set at the baseline $+10\%$ of the observed vocal-induced envelope amplitude range (10%-of-range). For example, if the baseline was 30 mV, and the maximum EMG envelope was 500 mV, then the threshold was set at $30\ \mathrm{mV}+(500-30)\ \mathrm{mV}/10 = 77\ \mathrm{mV}$, with an accuracy of 2 mV. Low-pitch vowel production was chosen to elicit the maximum EMG activity from the neck strap muscles. The value of 10% of the range was based on informal tests with two subjects during preliminary sessions. It was observed that the 5% level produced too much involuntary triggering of the EMG-EL, while the 20% made it hard to maintain voicing for the duration of a full sentence.

Within each trial, a sequence of visual commands with different background colors was displayed on the video monitor. The bottom panel of Fig. 3 shows a schematic of the sequence of events in a trial. Each trial started with a rest period of 10 s, followed by a "get ready" period of 1 or 2 s. The variable ready period was chosen to reduce the effect of the subject's anticipation and to minimize the reaction time [16]. The ready period was followed by the "response" cue, where the display showed the command "Say /a/" on a bright green background. The response interval was 2, 2.5, 3, or 3.5 s. Ending the response period was the stop signal, which was a display of the word "Stop" on a red background for 2 s. The stimulus sequence was repeated until all trials in a run were completed.

A run included 8 trials, each with a pseudo-randomized ready and response period length combination. Each voice source (Button-EL, EMG-EL, TE, and normal voice) was tested for five runs. A run was repeated only in the infrequent event of failure to maintain voicing during the response period. The order of testing of the different voice sources was also randomized for each of the normal subjects. There were $5\ \mathrm{runs} \times 8\ \mathrm{trials} \times 3\ \mathrm{voice\ sources} = 120\ \mathrm{tokens\ per\ subject}$, or 40 tokens per subject per voice source. The mean of the 40 tokens was used for statistical analysis, using a one-way repeated measures analysis of variance on the reaction times from the three different voice sources within the 7 normal subjects.

C. Threshold Optimization

The effect of threshold level was also evaluated using EMG signals obtained from neck strap muscles during normal voice production. The EMG signal was used to simulate the output of the EMG-EL at different threshold levels. Errors were calculated based on the discrepancy between the timing of the EMG-EL output and normal voice. An optimized threshold was found by error minimization, and was compared to the 10%-of-range threshold used during EMG-EL performance assessment.

EMG signals were recorded from ten normal subjects. The five male and five female subjects were native speakers of English, naive to the research, and had no history of neck surgery, voice disorders, or muscular pathology. The EMG recording setup was identical to that used for EMG-EL performance assessment, with the exception that the EMG signals were digitized rather than used to control the EMG-EL.

Neck surface EMG was recorded using a two-channel DelSys Bagnoli system (DelSys Inc., Boston, MA). Head-stage gain for all EMG electrodes was calibrated using SYSid software (Bell Laboratories). Rib cage and abdomen movements were recorded using a Respitrace System (Ambulatory Monitoring Inc., Ardsley, NY). A condenser microphone (Sony ECM-50PSW) was located 15 cm from the subject's mouth. Signals were recorded with a PC-based data acquisition system by Axon Instruments (Foster City, CA), that included computer controlled filters/amplifiers (CyberAmp models 320 and 380) connected to a 16-bit a/d board with up to 16 input channels (DigiData 1200). EMG signals were bandpass filtered at 10–2000 Hz during acquisition. All signals were digitized at 20 kHz sampling rate.

The subjects were asked to perform several tasks including sustained vowels (1–3 s long) at varying levels of loudness and pitch, as well as other nonvocal behaviors that involved strap muscle activation (e.g., tongue retraction and head movements). Tasks were repeated a minimum of three times. The data files from the vowel production task provided a speech signal and an EMG signal. Previous studies showed that that low pitch vocalization produced larger EMG amplitudes than at high pitch [17]. The amplitudes of the EMG envelope peaks were compared across the five different voicing conditions to confirm that low pitch vowel production yielded the most robust EMG signals. Consequently, the low pitch vowels were utilized for finding the optimum thresholds.

The EMG signal was used to determine when the EMG-EL device would have been turned on and off, and how that compared to when normal voice was being produced. The envelope of the EMG signal was tracked using a software-based third-order digital Butterworth LPF with 5-Hz cutoff frequency, mimicking the fast envelope of the EMG-EL processor. The resulting envelope was used to simulate the behavior of the

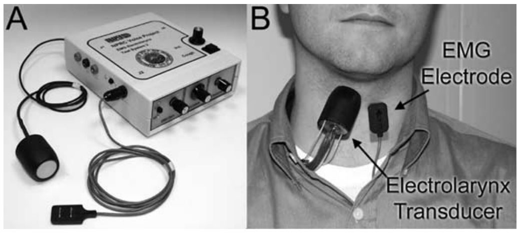

Fig. 4. (A) The EMG-EL System, showing DelSys surface EMG electrode, the EMG-EL processor, and EL transducer. (B) The transducer and EMG electrode are worn on the neck for hands-free EL speech.

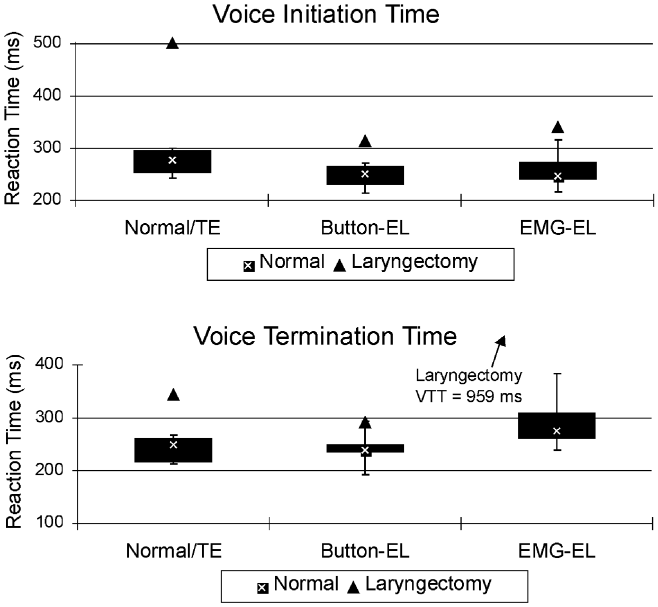

Fig. 5. Box plots showing the median, inter-quartile, and min-max values of voice initiation and termination times for 7 normal subjects using normal voice, button-EL, and EMG-EL. Solid triangles show reaction times of a laryngectomy subject with TE voice, button-EL and the EMG-EL.

EMG-EL processor by determining the times at which it would turn the transducer on and off given a constant threshold. The optimum threshold level was determined by minimizing the discrepancy between actual voicing and software-simulated EMG-EL voicing.

Two types of errors were defined: type I errors for when the EMG-EL was on although the subject was silent, and type II errors for when the EMG-EL was off although the subject was voicing. For analysis, type I and type II error durations were measured for each token. In order to compensate for the fact that each data file contained voicing and silent times of different lengths, the type I error duration was normalized by the total time within the data file when the subject was silent, and the type II error duration was normalized by the total voicing time. Therefore, type I and type II errors both ranged from 0 to 100 percent. After calculating the errors in each data file, the two error types were averaged over all repetitions of the task, which normally produced an $\mathrm{N} = 9$ for each subject (saying /i/, /a/, and /u/ three times each at low pitch). An error function was constructed by plotting the total amount of error (type I + type II) at each threshold level. The optimum threshold was defined as the point at which the total error was at a minimum. In order to compare the optimum thresholds across various subjects, the baseline activity of the muscle was used as a normalization factor. Therefore, the thresholds were calculated as percent increase above baseline envelope amplitude, where the baseline was subtracted from the threshold, then divided by the baseline and multiplied by 100%. For comparison, a second set of thresholds was calculated using the same data files with the fixed 10%-of-range method used during EMG-EL performance assessment.

III. Results

A. The EMG-EL System

The EMG-EL processor prototype was built and optimized for battery power consumption in collaboration with Draper Labs (Cambridge, MA). The processor was designed for use in conjunction with a DelSys DE2.1 electrode and a neck-mounted sound transducer. The dimensions of the processing unit are 5 cm × 15 cm × 17.5 cm, with an 1/8 in. mono jack connector for the EL transducer, a hypertronics 4-pole connector for the DelSys EMG electrode, and 5 knobs for setting threshold, EMG gain, starting fundamental frequency, transducer volume, and LPF corner frequency [Fig. 4(a)].

The DE2.1 DelSys surface EMG electrode was 2 cm × 4.1 cm × 0.5 cm in size, with two silver bars 1 cm long and 0.1 cm wide, spaced 1 cm apart. The electrode required a supply voltage of ±4.5 V, with 2.5-mA maximum supply current, and provided a head gain of $10 \pm 5\%\ \mathrm{V/V}$. The EL sound transducer used in conjunction with the EMG-EL system was identical to that used in NuVois (Mountain Precision Mfg., ID) EL devices. It contained a magnet, coil, piston, and drum assembly. It was mounted on a metal brace made of a thick copper wire covered with plastic tubing for comfort. The brace was wrapped around the back of the neck and down the middle of the chest for support. The neck-brace was discreet once placed under a shirt [Fig. 4(b)].

The processor box containing the battery pack and electrical circuit was mounted on the user's belt. The physical design of the EMG-EL processor was largely determined by the batteries (9-Vtype for circuit and six AA type for transducer). The EMG electrode and processing circuit were composed of low-power analog components that required an average of 5 mA of current. Given that an alkaline 9-V battery typically has 500 mAh capacity [18], the electrode and processor could operate for 100 h before requiring battery replacement.

The transducer power requirement is much larger than that of the control circuit. The resistance of the transducer coil is approximately 10 Ohms, which implied that when the full 9 V of the battery were applied with a 50% duty cycle square driving waveform, the average current load was 450 mA. If we assume the capacity of each AA battery to be 1000 mAh at 450 mA average load [18], then the six AA alkaline batteries were estimated to function for approximately 13 h of continuous operation.

B. EMG-EL Performance Assessment

Fig. 5 shows the VIT and VTT averages for all subjects. The VIT of normal subjects was not significantly different across voice sources ($\mathrm{p} = 0.133$, $\mathrm{F} = 2.4$, and $\mathrm{N} = 7$). Therefore, VIT of the EMG-EL was not different from normal

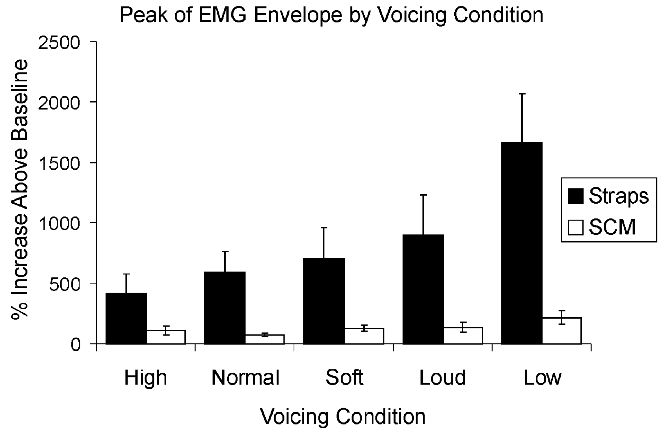

Fig. 6. Peak EMG activity level (avg. ± standard error) from the neck strap muscles (straps), and the sternocleidomastoid muscle (SCM), while producing a vowel under various conditions. The low pitch condition clearly exceeds all others for the straps, producing a peak activity of 1665% above baseline.

voice or the button-EL. For termination, the VTT averages across voice sources showed a significant overall difference ($p < 0.05$, $F = 5.5$, and $N = 7$). *Post-hoc* tests showed that the VTT of the EMG-EL was significantly larger than normal voice ($P < 0.05$), but not the button-EL. In the case of the laryngectomy subject, the VIT of the EMG-EL was 342 ms, which was larger than the 315 ms of the button-EL, but much smaller than the 503 ms of TE voice. EMG-EL VTT, on the other hand, was 959 ms, which is much larger than both the button-EL and TE voice.

Both normal subjects and the laryngectomy patient had the fastest VIT for the button-EL, followed by the EMG-EL VIT and then by the normal or TE VITs. For voice termination, the EMG-EL voice source was the slowest for both laryngectomy and normal subjects. Thus even though the laryngectomy subject tended to react more slowly than the normal subjects, the pattern of performance was similar across voice sources.

C. Threshold Optimization

The vowels produced at a low pitch were chosen as data for threshold optimization because the vocal-related EMG signals were much larger than other modes (see Fig. 6). Specifically, the low pitch vowels produced envelope amplitudes that averaged 1665% above baseline, which was approximately four times larger than those produced with high pitch vowels. This high SNR was specific to the midline strap muscles, as evidenced by simultaneous recordings from the sternocleidomastoid (SCM) muscles, which yielded EMG envelopes that were relatively much weaker, averaging 216% above baseline during low-pitch vowel production.

The EMG data was used to simulate the behavior of the EMG-EL at a given threshold, and that behavior was compared to normal voice in order to measure the amount of type I and type II errors. Fig. 7 shows one subject's cost function, calculated by adding type I and type II errors, and plotting the total as it varied with threshold. The optimum threshold was derived from the minimum point in the cost function. In this case, the optimum threshold was 259% above baseline, yielding type I errors 27% of the silent time, and type II errors only during 4% of the voicing time. The average optimum threshold from all 10 subjects was 273% above baseline, with a standard error of 103% above baseline. The average type I error probability across all subjects was 30%, while type II error probability was 11%.

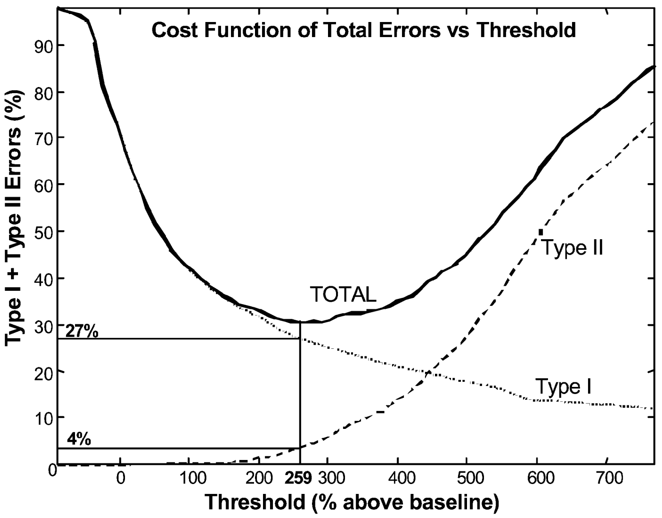

Fig. 7. Example of optimum threshold determined from the cost function constructed by adding type I (dotted) and type II (dashed) errors to form the total (solid) error. The optimum threshold is that which produces the minimum amount of total error.

When the 10%-of-range method was applied to the recorded EMG data for the ten additional normal subjects during low-pitch vowel production, the average threshold was 140%, which was not statistically different from the average optimum threshold (paired t-test, $p = 0.17$, $N = 10$). Therefore, the 10%-of-range method used for setting the EMG-EL threshold during performance testing produced thresholds that were statistically indistinguishable from the error optimization-derived thresholds.

IV. Discussion

A. The EMG-EL System

The main goal of using EMG signals to operate a neck-mounted EL voice device was to provide hands-free control. The use of EMG signals to control prosthetic devices has been well established, and was chosen for controlling EL speech due to the simplicity of the signal processing, as well as the portability of the physical hardware needed for its implementation. The prototype EMG-EL device is based on a simple processing scheme constructed from analog circuit components. The use of digital circuits could provide more sophisticated processing of EMG signals and thereby potentially improve the performance of the device. Nevertheless, the current method of processing appears to be sufficient as a proof of concept for using EMG signals to control an EL voice replacement device. Further improvements could be made to the physical design of the processor in order to reduce its size down to that of a pager worn on the belt. Additionally, battery-recharging capabilities could be added for efficiency and convenience with everyday use.

Possible alternative approaches to replace the use of the hand include mechanical and airflow based respiratory sensors. For example, one might attempt to detect the speaker's intention to start and stop speaking from chest motion (strain gauge) or stoma airflow (thermo-sensitive resistor). The potential problem with such systems is that they would require discernable changes in the breathing pattern to indicate the start and end of phonation. However, laryngectomy subjects might find it difficult to modulate their breathing pattern because of the lack of control over airway resistance through the tracheo-stoma.

B. EMG-EL Performance Assessment

Subjects using their normal voice produced VIT values that averaged 274 ms, with a standard deviation of 25 ms. These results match the findings of previous reports very well [16], where the predictive equation for the VIT yields 283 and 266 ms at the 1- and the 2-s ready-period lengths, respectively. Normal subjects also produced an average VTT value of 240 ms, with a standard deviation of 25 ms, which is very close to the 255 ± 61 ms produced in previous work by normal subjects voicing for 1.5–4 s [13]. The close resemblance of the normal subjects' results with those collected by other investigators confirms the validity of the setup and procedure used to collect the VIT and VTT data.

The VIT results showed that the EMG-EL is as fast as normal voice in initiation, and faster than TE voice in a laryngectomy subject. The slightly faster EL voice initiation is probably due to the fact that it takes less time to press a button than to build up subglottal pressure and adduct the vocal folds for normal voice production, or to occlude the stoma and push air through the TE valve for alaryngeal voice production. However, VIT of the EL was not significantly different than the other voice sources in normal subjects, and was only faintly smaller than the VIT of the EMG-EL used by the laryngectomy subject.

The finding that the EMG-EL voice initiation was not significantly different than normal or EL voice in normal subjects might be explained by the fact that neck strap muscle EMG leads voice production in timing. Previous studies have shown that the sternohyoid and sternothyroid strap muscles were found to produce EMG activity that precedes voice by 120 and 70 ms, respectively [17]. Similarly, the laryngectomy subject's EMG-EL voice initiation may have been faster than TE voice because laryngeal nerve activity precedes vocal fold movement by hundreds of milliseconds [19]. It was also found that TE voice initiation was slower than normal voice initiation, which might be explained partly by the older age of the laryngectomy subject, and partly by the additional time required to occlude the stoma, build up tracheal pressure, and open the TE valve before being able to initiate esophageal vibration to produce alaryngeal voice.

The VTT results showed that EMG-EL voice termination was not as fast as the other voice sources tested, especially in the laryngectomy subject. Slow VTT of the EMG-EL might be explained by the fact that it depends on strap muscle relaxation, which is not an active process. Normal voice termination is achieved either by the active contraction of the laryngeal adductor muscles to completely close the glottis, or the action of the abductor muscles to pull the vocal folds apart. The active process of lifting the finger away from either the stoma or the button-switch produced speedy TE and button-EL voice termination. The EMG-EL scheme, however, lacked a corresponding active mechanism that could terminate voicing as quickly. Furthermore, the low-pass filtering of the EMG signal and the relatively low threshold setting of 10%-of-range above the baseline produced fast VITs at the expense of slow voice termination.

These issues could be addressed through the use of different thresholds for voice initiation and termination, or the utilization of a second muscle EMG source that would act as an active EMG-EL voice terminator. In addition, training and practice might yield better conscious control over the neck muscles triggering the EMG-EL.

In addition to being slower than all other voice sources, the laryngectomy subject's EMG-EL VTT was found to be much larger than that of the normal subjects. This finding might be due to the laryngeal nerve supply controlling the laryngectomy subject's strap muscles. An important characteristic of the intrinsic laryngeal muscles is that some of them were found to exhibit a burst of activity after the termination of voicing [19]. Therefore, this post-phonatory activity might be largely responsible for the laryngectomy subject's lingering EMG-EL phonation after the stop signal.

Although subjects were unable to produce an EMG-EL VTT comparable to that obtained with normal and TE voice, this result might not have a serious impact when the EMG-EL is used for speech. During speech production, the user can possibly anticipate when pauses will occur with experience and training, and, therefore, adjust the activity of the muscles controlling the device to produce timely voice termination. Finally, the VIT and VTT are only two of many metrics that could be used to evaluate the efficacy of EMG-EL speech. Future work will be directed at testing the EMG-EL during meaningful vocal communication including production of words, sentences, and paragraphs. In addition, preliminary observations indicate that natural-sounding intonation can be readily incorporated into EMG-EL voice by using the slow EMG envelope to modulate the fundamental frequency.

C. Threshold Optimization

The process of finding the optimum thresholds was performed using recorded EMG and voice signals. Unlike the reaction time experiments, there was no feedback for the subjects, which could have influenced neck strap muscle activity and the resulting EMG-EL behavior. The 10%-of-range method of setting the threshold during EMG-EL performance testing was formulated using the EMG-EL device with real-time feedback. During pilot testing, two normal subjects listened to the output of the EMG-EL and adjusted the threshold according to the sensitivity they deemed appropriate. Interestingly, comparison of the optimum and the 10%-of-range thresholds showed no significant differences.

The thresholds from the low pitch vowel 10%-of-range method averaged 140% above baseline, while the optimum thresholds averaged 273%. This indicates that thresholds set according to subjective assessment of EMG-EL performance are lower than thresholds calculated by balancing type I and type II errors. Thus, it remains unclear what relative impact false triggering (type I) and nontriggering (type II) errors actually have on perceived EMG-EL speech. Adjusting the relative weights between the two types of errors during the optimization process produced different optimum thresholds. More specifically, increasing the relative cost of type I to type II errors produced larger optimum thresholds, and vice versa. When the relative cost of type I to type II weights were adjusted for each subject to make the optimum threshold match the 10%-of-range threshold, the average relative cost ratio across all ten subjects was 0.66. This implied that when choosing the 10%-of-range above baseline scheme, users of the EMG-EL generally preferred type I errors over type II errors. In other words, users provided with feedback from the behavior of the EMG-EL device opted to set the threshold such that unintended triggering of voice was less important than interruptions during intended voicing.

The average probabilities of error at the optimum threshold were 30% and 11% for type I and type II, respectively. This implied that during the time when the user intended to be silent, the device would have turned on 30% of the time, whereas if the user attempted to produce voice, the device would have turned on 89% of the time. Employing different initiation and termination thresholds rather than using a fixed threshold level could likely have reduced these error levels. Furthermore, training the users to better control their EMG activity might reduce these errors as well.

V. Conclusion

This paper presents evidence that EMG signals from neck strap muscles can be used effectively to control the initiation and termination of electrolarynx voice. The main advantage of the EMG-EL system is its hands-free control, compared to EL and TE speech where the use of one hand is normally required. The results from the voice initiation data showed that the EMG-EL is as good as normal voice and a commercial EL device, and faster than TE voice. Normal subjects were also able to achieve EMG-EL voice termination that was not different than that produced with a commercial EL. Future work will focus on examining the role that formal training may play in optimizing EMG-EL use, coupled with the evaluation of EMG-EL performance during meaningful vocal speech communication.

Acknowledgment

E. A. Goldstein would like to dedicate this paper in memory of Dr. T. McMahon, who provided the initial guidance, support, and inspiration for this project.

References

[1] R. E. Hillman, M. J. Walsh, G. T. Wolf, S. G. Fisher, and W. K. Hong, "Functional outcomes following treatment for advanced laryngeal cancer," *Ann Otol Rhinol Laryngol Suppl*, vol. 172, pp. 1–27, 1998.

[2] S. Gray and H. Konrad, "Laryngectomy: Postsurgical rehabilitation of communication," *Arch Phys Med Rehabil*, vol. 57, no. 3, pp. 140–142, 1976.

[3] 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, no. 6, pp. 503–510, 1992.

[4] R. E. Hillman *et al.*, Department of Veterans Affairs Rehabilitation Research and Development Grant # C2243–2DC, 1998.

[5] S. C. Jacobsen, D. F. Knutti, R. T. Johnson, and H. H. Sears, "Development of the Utah artificial arm," *IEEE Trans Biomed Eng.*, vol. 29, no. 4, pp. 249–269, 1982.

[6] G. N. Saridis and T. P. Gootee, "EMG pattern analysis and classification for a prosthetic arm," *IEEE Trans Biomed Eng.*, vol. 29, no. 6, pp. 403–12, 1982.

[7] A. Latwesen and P. E. Patterson, "Identification of lower arm motions using the EMG signals of shoulder muscles," *Med Eng Phys.*, vol. 16, pp. 113–121, 1994.

[8] M. Yamada, N. Niwa, and A. Uchiyama, "Evaluation of a multifunctional hand prosthesis system using EMG controlled animation," *IEEE Trans Biomed Eng.*, vol. 30, no. 11, pp. 759–63, 1983.

[9] Y. Koike and M. Kawato, "Estimation of dynamic joint torques and trajectory formation from surface electromyography signals using a neural network model," *Biol Cybern*, vol. 73, pp. 291–300, 1995.

[10] A. Junker, "Brain-Body Actuated System," United States Patent, 5 474 082, Dec. 12, 1995.

[11] S. Scargle, "EMG/EEG Head-Computer-Interface System for Computer Cursor Control," M.Sc., International University of Florida, 1998.

[12] M. Cheng, G. Xiaorong, G. Shangkai, and X. Dingfeng, "Design and implementation of a brain-computer interface with high transfer rates," *IEEE Trans. Biomed. Eng.*, vol. 49, no. 10, pp. 1181–86, 2002.

[13] W. L. Cullinan and M. T. Springer, "Voice initiation and termination times in stuttering and nonstuttering children," *J. Speech & Hear. Res.*, vol. 23, pp. 344–360, 1980.

[14] M. R. Adams and P. Hayden, "The ability of stutterers and nonstutterers to initiate and terminate phonation during production of an isolated vowel," *J. Speech & Hear. Res.*, vol. 19, pp. 290–296, 1976.

[15] T. Shipp, K. Izdebski, and P. Morrissey, "Physiological stages of vocal reaction time," *J. Speech & Hear. Res.*, vol. 27, pp. 173–178, 1984.

[16] B. Watson, "Foreperiod duration, range, and ordering effects on acoustic LRT in normal speakers," *J. Voice*, vol. 8, no. 3, pp. 248–254, 1994.

[17] J. Atkinson, "Correlation analysis of the physiological factors controlling fundamental voice frequency," *J. Acoust. Soc. Am.*, vol. 63, no. 1, pp. 211–222, 1978.

[18] P. Horowitz and W. Hill, *The Art of Electronics*, 2nd ed: Cambridge University Press, 1995.

[19] A. D. Hillel, "The study of laryngeal muscle activity in normal human subjects and in patients with laryngeal dystonia using multiple fine-wire electromyography," *Laryngoscope*, pt. Part 2, vol. 111, no. 4, 2001.

Ehab A. Goldstein was born in Amman, Jordan, in 1976 and received the B.S. degree *summa cum laude* in biomedical engineering from Harvard University, Cambridge, MA, in 1998. He received the Ph.D. degree in biomedical engineering from the Division of Engineering and Applied Sciences at Harvard University in November 2003 . He also received a certificate for the completion of the requirements for the Ph.D. degree in speech and hearing biosciences and technology at the Harvard-MIT division of Health Sciences and Technology. His thesis is titled "Prosthetic Voice Controlled By Muscle Electromyographic Signals'" and is focused on the design and implementation of an improved alaryngeal voice prosthesis system based on the use of EMG signals from the infrahyoid neck strap muscles.)

James T. Heaton was born in Wilmington, DE, in 1967. He received the B.A. degree in psychology from Luther College, Decorah, IA, in 1990, and the M.S. and Ph.D. degrees in psychology from the University of Maryland, College Park, MD, in 1993 and 1997, respectively.

He joined the faculty of the Department of Otology and Laryngology, Harvard Medical School, Cambridge, MA, in 1997 where he is currently an Assistant Professor and a Fellow of the 50th Anniversary Program for Scholars in Medicine. He also holds adjunct faculty positions in the Department of Communication Sciences and Disorders at both the Massachusetts General Hospital Institute of Health Professions and Emerson College, Boston, MA, where he teaches anatomy and physiology to graduate students seeking the M.S. degree in speech language pathology. His research interests include the neural mechanisms of vocal learning and vocal production in birds and mammals. His recent research projects have focused on human larynx innervation (neurolaryngology), exploring new methods for quantitative assessment of laryngeal function as well as methods for establishing a link between laryngeal nerves and prosthetic equipment for neural control of an artificial voice source.

Dr. Heaton is a member of the Society for Neuroscience.

James B. Kobler received the B.A. degree in biology from Vassar College, Poughkeepsie, NY, an 1975 and the Ph.D. degree in neuroscience from the University of North Carolina at Chapel Hill in 1983.

He was a postdoctoral fellow in the Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, from 1983–1986. Since 1991, he has been the Director of the Harris Peyton Mosher Laryngological Research Laboratory in the Department of Otolaryngology, Massachusetts Eye and Ear Infirmary, and a member of the Department of Otology and Laryngology, Harvard Medical School. He is also an active member of the Harvard-MIT Division of Health Sciences and Technology. His main research interests currently are in the areas of speech production, laryngeal physiology and development of instrumentation for research, and clinical applications in laryngology.

Garrett B. Stanley (S'96–A'97) received the B.M.E. degree with highest honors from the Georgia Institute of Technology, Atlanta, in 1992, and the M.S. and Ph.D. degrees in mechanical engineering from the University of California at Berkeley in 1995 and 1997, respectively.

From 1995 to 1997, he was an American Heart Association Predoctoral Fellow. From 1997 to 1999, he was a Postdoctoral Fellow in the Neuroscience Division of the Department of Molecular and Cell Biology at the University of California at Berkeley, and an NIH Postdoctoral Fellow. He is currently an Assistant Professor of Biomedical Engineering with the Division of Engineering and Applied Sciences at Harvard University, Cambridge, MA, and is an active member of the Harvard-MIT Division of Health Sciences and Technology (HST). His research interests include biological signal processing, experimental and theoretical approaches for understanding neural coding in sensory systems, neuronal point processes, parameter estimation, and the development of devices for recording from and stimulating the nervous system.

Robert E. Hillman received the B.S. and M.S. degrees in speech-language pathology from the Pennsylvania State University, University Park, in 1974 and 1975, respectively, and the Ph.D. degree in speech science from Purdue University, West Lafayette, IN, in 1980

He is a Fellow of the American Speech-Language-Hearing Association and currently holds positions as Director of Clinical and Research Programs in Speech Pathology and Director of the Voice and Speech Laboratory at the Massachusetts Eye and Ear Infirmary, Associate Professor in Otology and Laryngology at Harvard Medical School, Professor in Communication Sciences and Disorders at the Massachusetts General Hospital Institute of Health Professions, and a Research Affiliate in the Speech Communication Group at MIT's Research Laboratory of Electronics. He has been awarded over 18 grants from governmental and private sources to support his research, with funding from the National Institutes of Health (NIH) since 1984. His research and numerous publications have focused on mechanisms for normal and disordered voice production, evaluation and development of methods for alaryngeal (laryngectomy) speech rehabilitation, development of objective physiologic and acoustic measures of voice and speech production, and evaluation of methods used to treat voice disorders.