Using visual feedback to enhance intonation control with a variable pitch electrolarynx

Noor Al-Zanoon, Vijay Parsa, and Philip C. Doyle

Citation: The Journal of the Acoustical Society of America **147**, 1802 (2020); doi: 10.1121/10.0000936
View online: https://doi.org/10.1121/10.0000936
View Table of Contents: https://asa.scitation.org/toc/jas/147/3
Published by the Acoustical Society of America

JASA ARTICLE

Using visual feedback to enhance intonation control with a variable pitch electrolarynx

Noor Al-Zanoon,[1,a)] Vijay Parsa,[2,b)] and Philip C. Doyle[2,c)]
[1]*Department of Communication Sciences and Disorders, University of Alberta, 116 Street and 85 Avenue, Edmonton, Alberta T6G 2R3, Canada*
[2]*School of Communication Sciences and Disorders, Elborn College, Western University, London, Ontario N6A 3K7, Canada*

ABSTRACT:
This study evaluated the effectiveness of using visual feedback to facilitate pitch control by a speaker using a pressure sensitive onset controlled electrolarynx (EL). This proof-of-concept study was conducted with one healthy adult. The participant-speaker was provided with computer generated visual feedback over five sessions within a consecutive period of three weeks. Changes in force control accuracy were gathered and analyzed. An improvement in finger (thumb) force control accuracy from the first to the last training session was documented. The results of this study provide data toward the development of a clinical training protocol for the use of a pressure sensitive onset controlled EL by laryngectomized speakers. Further, these results highlight the importance of developing a relevant multimodality training protocol for the improvement of postlaryngectomy EL speech production.
 https://doi.org/10.1121/10.0000936

(Received 25 September 2019; revised 28 February 2020; accepted 3 March 2020; published online 20 March 2020)
[Editor: Anders Lofqvist] Pages: 1802–1811

I. INTRODUCTION

Total laryngectomy (TL) is a surgical procedure used for the treatment of laryngeal cancer. This procedure results in the removal of the entire larynx, including the vocal folds. After a TL, normal voice production is lost and a new, alaryngeal vibratory source must be acquired for the generation of voice and speech. One common postlaryngectomy communication method involves the use of an electronic artificial larynx or what is commonly termed an "electrolarynx" (EL). While recent estimates of the percentage of laryngectomees who use an EL are unavailable, previous reports have shown that more than half use an EL as their primary postlaryngectomy communication method, or as a back-up method of communication (Hillman *et al.*, 1998).

An EL is a hand-held, battery operated device that provides an external voicing source that can be articulated into speech (Doyle *et al.*, 2005). These EL excitations are either transmitted into the oral cavity by a tube placed within the mouth (transoral devices) or more commonly through the neck tissues (transcervical devices). All EL devices are designed with a manual control button that allows an EL user to actively control signal onset and offset when speaking. Over the past decade, some ELs have been manufactured with an additional onset button or one with a digitally optimized potentiometer that permits active and dynamic manipulation of the source frequencies. That is, some newer EL devices can generate a smooth range of frequency in response to varied finger pressure during device activation. This feature and its potential reduction in a monotonous signal source enhance the overall effectiveness of the EL speech. However, despite this design feature, one of the most problematic aspects of training focuses on the speaker's ability to actively and appropriately vary pitch. This lack of pitch variation results in EL speech being judged by listeners to be monotone and robotic (Bennett and Weinberg, 1973; Ma *et al.*, 1999; Meltzner and Hillman, 2005; Saikachi *et al.*, 2009; Tanaka *et al.*, 2014; Uemi *et al.*, 1994).

The role of frequency variation or "intonation" in English verbal speech serves multiple functions, including the speaker's ability to (1) express emotional states, (2) distinguish between different types of utterances (e.g., questions and declarative statements), (3) emphasize key words in utterances (Vaissière, 2005), as well as (4) provide lexical and syntactic distinctions between nouns/verbs (Fodor and Garrett, 1967). These factors collectively create a melodic, frequency modulated signal that allows for enhanced verbal communication. EL speech, however, has been shown to exhibit significant signal restrictions that impact the effectiveness of communication; this includes a reduction in pitch variation, inability to generate prosodic contours, and low frequency signal deficits (Barney, 1958) with a collective influence on speech quality and intelligibility (Miller *et al.*, 2010; Weiss *et al.*, 1979).

Early studies have reported an improvement in the quality of EL speech with variation of intonation. For example,

a)Electronic mail: alzanoon@ualberta.ca
b)Also at: Department of Electrical and Computer Engineering, Thompson Engineering Building, Western University, London, Ontario N6A 5B9, Canada.
c)Also at: Department of Otolaryngology Head and Neck Surgery, Stanford University School of Medicine, Stanford, CA 94035, USA.

Watson and Schlauch (2009) found that sentences produced using an EL with a variable F_0, compared to the same sentences with acoustically flattened F_0, were better understood by listeners. Similarly, Gandour *et al.* (1980) reported that listeners struggled to differentiate between sentences (declarative and interrogative) produced without a variable F_0. These studies highlight the importance of producing dynamic pitch fluctuations required for the signaling of linguistically meaningful contrasts to the listener (Gandour *et al.*, 1980; Gandour and Weinberg, 1982, 1984).

As a result of the noted frequency limitations, several studies have attempted to improve the design of the EL by including a dynamic and speaker-activated pitch control option. Theoretically, dynamic pitch control would allow an EL speaker to produce changes in pitch throughout a conversation with assumed enhancements in communication effectiveness (e.g., Liu *et al.*, 2006; Meltzner *et al.*, 2001). To date, two methods of speaker-activated pitch control have been researched: hands free and manual control (using a potentiometer). Goldstein *et al.* (2007) used electromyographic signals from the neck to modulate EL pitch control. Goldstein *et al.* (2007) reported that all participants were able to successfully control pitch using the hands-free EMG controlled EL. Despite this success rate, the authors reported difficulty in starting and stopping voicing using EMG control. The difficulty in controlling voicing via the use of EMG stems from the unknown role of neck muscle relaxation. Another method of producing variable F_0 is facilitated via the use of finger-controlled activation. Using this method, pitch is modulated by an increase or decrease of pressure on a single on-off control button. In this situation, finger pressure is detected and modified by a force potentiometer (Liu and Ng, 2007). Similar to the limitations found in EMG controlled ELs, finger-controlled pitch remains a problem for many EL users (Goldstein *et al.*, 2007).

Given the increased prominence today of EL devices that offer the feature of dynamic finger pitch control, questions about the speaker's capacity to directly and manually effect such a change have emerged. Yet it is critical to acknowledge that while the instrumental potential to modulate frequency does exist for some EL devices, the user's ability to volitionally manipulate a finger-based control to effect such a change is not automatic as it requires complex coordination. Given previous literature which has shown that individuals have difficulty using a pressure sensitive EL device to control intonation, we believe that the acquisition of intonation when using a finger controlled EL is a complex motor learning problem that deserves empirical evaluation.

One factor shown to enhance motor learning is the provision of direct "feedback" or information given to an individual about the motor task they seek to acquire. The type of information provided to the individual can be divided into two categories: knowledge of results (KR) and knowledge of performance (KP). KR is information related to the outcome of a participant's performance (Magill and Anderson, 2007). KP is information related to specific characteristics of the motor movement associated with the learned task (Magill and Anderson, 2007). Both KR and KP are then delivered through different mediums (e.g., verbal, visual, and proprioceptive).

Delivery of KR and KP through visual feedback has been shown to enhance learning of a variety of motor tasks (e.g., Lee *et al.*, 1990; Snodgrass *et al.*, 2010; Ossmy and Mukamel, 2018). Several researchers report the use of visual feedback in learning complex force production tasks. For example, Therrien and Balasubramaniam (2010) used a force transducer—a small device that measures the pinch grip force between fingers. This device was used to measure participant responses to specific instructions to press a sensor with a specified manual force. Participants who did not receive visual feedback on the force they applied tended to overestimate or underestimate the amount of force needed to satisfy the trial instructions. In contrast, participants who were provided with online visual feedback were able to respond with the appropriate amount of force. Therefore, we conducted a proof-of-concept study designed to systematically develop and test a training protocol to facilitate the acquisition of intonation using a pressure sensitive EL. More specifically, this study addressed the following question: will the use of real-time, online visual feedback enhance the finger pressure control in manual EL devices?

II. METHODS

A. Participant

The participant-speaker was a healthy 67.0-year-old, right-handed male, who identified as a monolingual native English speaker and reported no history of speech, language, or hearing deficits. The participant was recruited through the professional contacts of one of the authors. The participant-speaker was considered naive as he had no prior exposure or knowledge related to laryngectomy or alaryngeal speech or to an EL device, nor did he have any knowledge as to its use. Additionally, he was not briefed about the primary purpose of the training paradigm at any point during the experiment. The participant-speaker participated in four, one-hour sessions that spanned the period of three weeks. Because males represent the largest proportion of those diagnosed with laryngeal cancer, we selected a typical speaking, healthy 67-year-old male to serve as our experimental participant. Further, because this project was designed as a proof-of-concept study, we deliberately sought to construct an experimental paradigm that would reduce other potential motor challenges that might coexist in those who have undergone laryngectomy.

B. Speech Stimuli

1. Design parameters

Speech stimuli specifically created for this study included echoic questions and declarative pairs. Echoic questions are sentences that are identical to their declarative counterparts; however, they differ in the F_0 characteristics that occur at the end of the sentence. For example, the following sentences form an echoic question and declarative statement pair: "Joe ate the soup!" and "Joe ate the soup?" The echoic question only differs from the declarative sample by the rise in the final F_0 on the word "soup" (Lieberman, 1967). This rise in terminal F_0 and its contour permits coding of the sentence as an interrogative, rather than as a declarative. Four echoic question and statement pairs constructed to meet specific phonetic parameters and properties were created in order to specifically isolate F_0 in EL speech. These parameters included speech rate, word and level sentence stress, phonemes, and syntax. These experimental stimuli were designed to include sentences that can be spoken in a single breath group by a normal speaker (i.e., no sentences had syntax that required a comma or semi-colon that would require even brief linguistic pausing). This was based on work by Rothman (1978) who found that poor EL users paused more often between phrase groups. Further, to control for word level stress, we constructed experimental sentences to be comprised of a majority of monosyllabic words. Finally, previous research has demonstrated that EL speakers have difficulty producing voiceless stop plosives (e.g., /p/) and affricates (e.g., /ʃ/) (Weiss and Basili, 1985; Weiss *et al.*, 1979). The final stimuli set can be found in Table I.

2. Recording and preliminary testing

The proprietary experimental stimuli were initially recited orally and recorded by two adult, native English speakers (1 male and 1 female) with a central Canadian dialect. All recordings were acquired in a professional sound-treated recording suite housed in the Voice Production and Perception Laboratory, Elborn College, at the University of Western Ontario. A head-set microphone (Shure SM10A) and a preamplifier/digitizer (M-AUDIO ProFire 610, 24 bit/ 192 kHz) were used for all recordings. Recordings were sampled at 44.1 kHz. The headset microphone was adjusted to an optimal distance from the corner of the participant-speaker's mouth. This distance was selected by having the participant speak into the microphone while the researcher observed the output on Audacity 2.1.2. The researcher then adjusted the distance based on the quality of the output (e.g., clipping). Audacity 2.1.2 was used to record and save all experimental samples, resulting in a stimulus set containing 92 samples (2 speakers × 23 sentences × 2 frequency profiles). Eight of the 92 samples were randomly selected and duplicated and used for reliability purposes.

TABLE I. The four echoic sentence pairs that comprised the experimental stimuli.

Declarative	Interrogative
Anna loves orzo.	Anna loves orzo?
Lane loves whales.	Lane loves whales?
Mary's mole was vile	Mary's mole was vile?
Myles knows Lane is wise.	Myles knows Lane is wise?

Once recorded, these stimuli were validated by 15 naive, normal-hearing English listeners using a forced choice listening paradigm. Participant responses were coded using a simple 1 or 0 coding system. If the stimulus item was correctly identified as either a question or a declarative statement, the participant listener received a score of 1. If incorrectly identified, they were given a score of 0. For each stimulus item, the number of correct identifications by listeners was added to generate a total score out of 92. Any stimulus item that received a score below 90% was removed from use in the training portion of the experiment. This validation task was performed to confirm that the stimuli met the stated F_0 properties they were developed to achieve prior to commencing the training protocol. This resulted in the four proprietary sentence and interrogative pairs.

3. Visual target creation

The four previously validated echoic pairs were used to create the computerized visual target bars necessary for training. To create these target bars, the force associated with the production of a specific F_0 was first determined. This was done by first setting and calibrating a force sensor and then associating force with recorded frequency. This procedure is described in detail in a subsequent section (see "force frequency relationship").

4. Sensor setting and calibration

One FlexiForce™ A201 sensor (Tekscan, Inc., Boston, Massachusetts) was placed on the pressure sensitive button of the TruTone™ EL. The Arduino Uno microcontroller board was used to sample and quantize the force sensor output. The digitized data was fed through a USB interface to a custom written MATLAB® (R2011a) (MATLAB, 2011) script running on a PC (Fig. 1). To calibrate the sensor, a 5-point calibration plot was generated. The generation of this plot was done using the recommended calibration sequence provided by the manufacturer of the force transducer (Tekscan, Inc.).

5. Force-frequency relationship

To create the visual target displays necessary for training, the force required to a produce a desired F_0 with the EL was identified. This was determined by applying a known force to the on/off button on the EL and recording the associated F_0. Recordings were obtained using a microphone (AKG C4000 B Condenser), preamplifier (MAUDIO, ProFire 610, 24 bit/ 192 kHz), and Audacity 2.1.2 software. To measure force during each recording session, one FlexiForce™ A201 sensor was placed on the pressure sensitive button of the EL and held at a specific force. As the pressure sensitive on/off button was maintained at a constant force, the audio was recorded using Audacity 2.1.2. The force was recorded at a sampling rate of 25 Hz, using MATLAB® through a script designed specifically for this experiment.

For each force pressure, an average was calculated in Excel over a period of 20 s. The duration of 20 s was arbitrary; however, we believed that a longer, rather that shorter

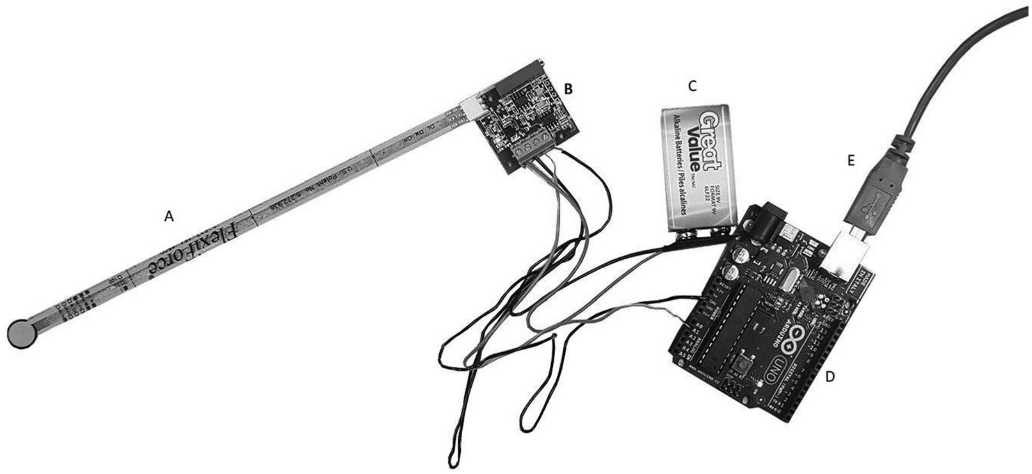

FIG. 1. (A) FlexiForce™ sensor, (B) FlexiForce™ Quickstart board, (C) 9 V battery, (D) Arduino Uno (LC066), (E) USB connection.

duration would allow for a more accurate determination of speaker performance during this complex task. A scatter plot with a line of best fit was then created and a linear equation was generated: $y = 0.0165x - 1.1667$ with an R^2 value of 0.8078. The y-value in this equation represents force (mV)[1] and the x-value represents F_0 (Hz). A positive linear relationship was found between force and frequency for the EL employed; that is, as finger force increased, F_0 (pitch) also increased. During the calibration sequence, the average standard deviation was found to be 8% over the 20 s period.

6. Conversion of frequency values

The F_0 values of the four echoic pairs were extracted using the "Get pitch" function in Praat 6.0.34 (http://www.fon.hum.uva.nl/praat/; Boersma, 2001). These frequency values were then converted into force values using the previously determined force-frequency relationship ($y = 0.0165x - 1.1667$). The predetermined force values were used to generate the specific target bars in the custom written MATLAB® script.

The visual target bars and stimuli were then used to create three training tasks: (1) Force Bars (FB) – visual target bars with no text, (2) Single Words (SW) – target bars with single words, and (3) Sentence (S) – target bars with full sentences. The target bars with single words were random words chosen from the previously validated echoic question and statement pairs, and the associated F_0 and force values were calculated using the same method as that performed for the full sentences.

7. Visual display

MATLAB® was used to display these force values on a force versus time graph for each stimulus. All stimuli were displayed on a computer with a 1050 by 1060 screen resolution. Target bars were pre-set with a thick line width to ensure clear visibility during the training protocol (time in seconds was represented on the *x* axis; force in mV was represented on the *y* axis).

C. Procedure

1. Experiment setup

Before positioning the EL and the force sensor, the participant was seated comfortably in front of a computer screen. The seat height and computer screen placement (75 cm away from the participant) were adjusted to ensure optimal viewing, and these positions were maintained in all sessions over the course of the experiment.

2. Placement of the EL

At the start of each training session, the optimal placement of the EL was identified in a systematic manner by the experimenter (Doyle, 1994). The primary researcher and an experienced clinician determined the best sound quality via both ear and EL point of contact. Once the optimal position on the participant's neck was identified, a customized silicone attachment with its center point open was placed at that location for the remainder of the training session. This attachment was used in order to maintain consistency of neck/device coupling during tasks. This attachment also permitted the EL to be removed from the neck as needed, but also insured that the identical placement and skin contact was maintained over time. The consistency of EL placement was of particular importance in eliminating potential confounds in the signal transmission through the participant's neck during the experimental tasks (Meltzner *et al.*, 2003).

3. Placement of the sensor

A specially designed Velcro harness was strapped onto the participant to allow for the placement of the sensor on the EL pressure sensitive button. The Arduino and FlexiForce™ Quickstart board were placed in separate plastic pockets lined with Velcro patches. Both pockets were then placed into the Velcro harness and positioned to allow the participant to hold down the FlexiForce™ sensor on the pressure sensitive button when coupled to the neck (Fig. 2). This set-up was used for all training tasks involving the EL

FIG. 2. Experimental set-up with the EL coupled to the neck.

coupled to the neck. For trials involving target bar matching without the EL coupling, the hardware set-up was placed on the desk. The FlexiForce^{TM} sensor was placed onto the EL pressure sensitive button. The participant could then hold both the sensor and the EL in his dominant hand without placing the apparatus on his neck or interfering with his movement.

4. Training sequence

Participant training took place in the same laboratory used in preliminary phases of the experiment. A clinician with more than 30 years of experience in postlaryngectomy voice and speech rehabilitation supervised the initial training session that involved the introduction and placement of the EL. Training that used visual feedback occurred over two consecutive weeks (two sessions per week), with each experimental session lasting approximately 1 h. After the two weeks of training was completed, the participant was evaluated during a separate 1-hr session in order to collect post-training data. Thus, the total experimental time consisted of three consecutive weeks.

Each session was divided into three phases: (1) review phase – the participant was given a quiz to review tasks learned and mastered in the previous session; (2) learning phase – a new task was introduced and practiced; and (3) preview phase – the participant practiced a skill to be learned the following week. Finally, all training tasks involved the placement of a sensor on the pressure sensitive button of the EL. During each session, the participant's force productions were displayed on a computer screen along with previously calculated target forces. Each session contained a total of 40 trials that were distributed across each of the phases: (1) 8 trials in the review phase (2) 24 in the learning phase, and (3) 8 trials in the preview phase. In the learning phase, the participant was given three attempts to learn each stimulus token.

5. Training tasks

a. Week 1 (Session 1). Before starting the training sequences, the researcher familiarized the participant-speaker with the linguistic function of intonation. After familiarization, the participant was instructed on the basic components of an EL and its function. Next, he was asked to complete baseline measures that included all three visual target tasks: FB task, SW task, and S task [see Fig. 3(a–d)]. Under these circumstances, the participant ultimately served as his own control throughout the course of the experiment with the trajectory of his EL skill acquisition determined empirically. Task performance throughout all sessions was rigorously monitored under strict experimental conditions. Additionally, at the beginning of any given session, the participant was always requested to return to the terminal task from the prior session. Based on our *a priori* notion that visual monitoring was essential to acquiring the force matching skill, this procedure appeared to reduce obvious potential confounds, given the training paradigm. While the participant did have the potential to identify the basic relationship between finger pressure and the target bar (in fact, that was the task instructed), there was no explicit mention of this task and its relationship to frequency change. However, during all tasks, the participant was able to simultaneously hear the EL source signal and, accordingly, it is likely that he was able to monitor and infer the relationship of increases and decreases in frequency during tasks.

b. Week 1 (Session 2). (1) FB task: The participant was shown the target force bar on a computer screen and asked to match the force using the pressure sensitive button on the EL [Fig. 3(a)]. Once the participant mastered controlling pressure on the EL without neck coupling, he was asked to place the EL on the neck in its natural position. The participant was then asked to practice the FB task with the EL coupled to his neck.

c. Week 2 (Session 1). During this session, the participant performed the SW task [see Fig. 3(b)]

d. Week 2 (Session 2). After the participant successfully performed the SW task, he was asked to practice the S task [Figs. 3(c) and 3(d)]. For all tasks involving speech production, the participant was asked to be as natural as possible for these spoken tasks; thus, we sought to eliminate potential confounds related to reductions in the participant's speech rate during these experimental tasks.

https://doi.org/10.1121/10.0000936

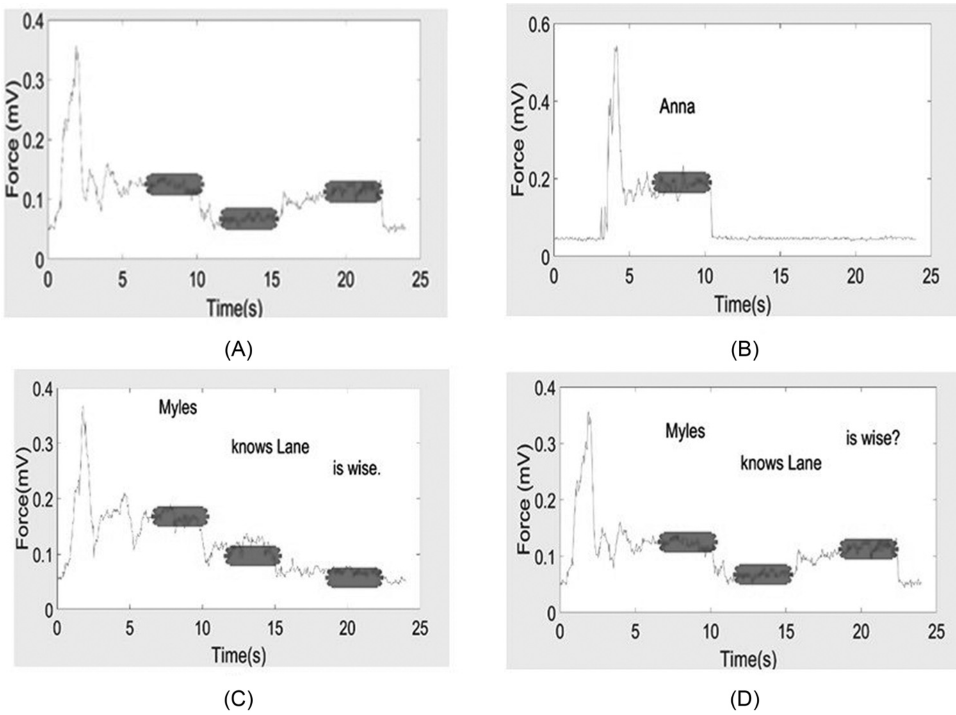

FIG. 3. Training Tasks: (A) FB task: Force bar matching task with no text; (B) SW task: Single words added to the force bar; (C) S task: Declarative added to the force target bars; and (D) S task: Question added to the force target bars. The initial large increase in force is the transient response resulting in "overshoot" as the subject initiated the control of the EL device.

e. Week 3 (Evaluation task). The participant was evaluated on all previously learned tasks (FB, SW, and S).

6. Evaluation criteria

To evaluate whether the participant was able to move from one newly learned skill to the next, the standard error (SE) between the target bar and force production was calculated and converted to a percentage. The acceptable error range per trial was set at 0.025 mV above and below the target bar. This range was experimentally determined by the research and engineering team to ensure that trial difficulty was set at an acceptable level. The SE was automatically calculated for each trial and was presented using MATLAB®. To pass from one trial to the next, the participant-speaker had to meet and/or exceed a force matching percentage (FMP) of 75%. This percentage was determined using trial and error assessment within the experimental protocol. Although the determination of 75% was selected arbitrarily, we believe it was an appropriate index for the confirmation of the participant-speaker's acquisition of the skill. Similarly, we believed this minimum level of performance ensured that the task was undertaken at an appropriate level of difficulty in order to document task mastery. If the participant's force productions fell within the predetermined SE, he could then advance to the next trial. If the participant's force productions were outside the predetermined SE, verbal feedback to adjust learning and conditions was provided. The participant was informed whether he passed or failed the trial immediately after its completion. This rendered the experimental design to include and provide both KR and KP. In the present study, KP is information on the thumb biomechanics, that is, the force at which the EL onset button was pressed. This was quantified as the FMP. The participant was not made aware of the FMP or the biomechanics of his thumb during the task. Rather, he was given general instructions to closely match the target bar to the best of his abilities. The participant was able to see the target bars, as well as their own force productions on the screen (see Fig. 3). Therefore, the participant did not receive KP feedback. However, the experimenter tracked and used KP as an objective method of deciding whether the participant could advance in the training protocol. KP was, therefore, used to quantify and operationalise the success of a trial, rather than subjective assessment of participant improvement. The frequency of visual feedback and KR remained identical over all training sessions.

D. Data analysis

The training paradigm included repeated measurements of the participant's force productions over the four sessions. For each session, a total mean and associated SE of the FMP values were calculated. For descriptive analysis, the mean and SE of the participant matched force values were plotted on a graph of time (session number) versus the average FMP values (Fig. 4).

To understand the effect of task type on FMP accuracy, an analysis of each of the three task types (FB, SW, and S) was also performed by calculating average per session and the associated SE of the PM value. These values were then

https://doi.org/10.1121/10.0000936

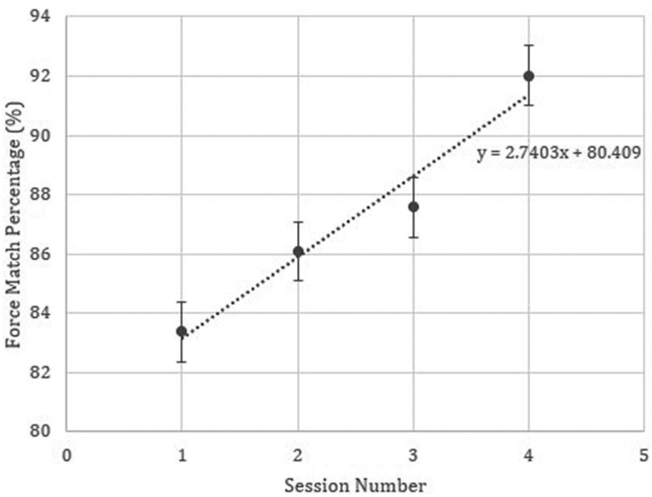

FIG. 4. Participant-speaker's FMP by training session. The participant's variability (mV) (standard error) from the average FMP is represented by the error bars.

plotted on a bar graph representing task type versus average FMP (Fig. 5).

III. RESULTS

A. Overall force matching performance

A positive linear relationship was found for the FMP over the four sessions. FMP across sessions was found to increase from the first to the final session with an improvement rate of 2.7403% in mV (Fig. 4). From the first session to the last session, the participant improved by 10.84% in force matching accuracy.

B. Force matching performance by task

The participant-speaker produced an average of 93.04% ($SD = 0.06$, $SE = 0.01$) on the FB task, followed by an average of 85.42% ($SD = 0.17$, $SE = 0.03$) in the SW production task. The S task was found to have the lowest FMP average of 76.01% ($SD = 0.14$, $SE = 0.03$) (Fig. 5). A relative decrease of 8.19% of matching was found between the FB and the SW task. A decrease of 11.01% was found between SW and S. Finally, a decrease of 18.30% was found between the FB and S task.

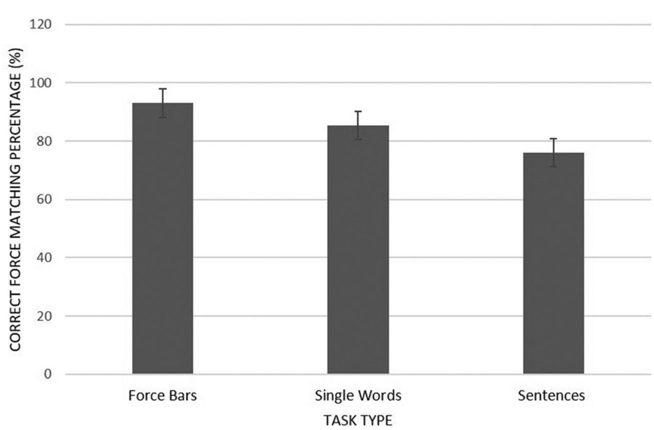

FIG. 5. Participant's performance on each of the three tasks. The error bars represent the overall force matching error (mV) across four sessions for each task.

IV. DISCUSSION

This project sought to empirically evaluate a novel sequential training paradigm designed to facilitate pitch control using a pressure sensitive EL device. We approached this finger control problem as a complex problem of coordinated events: (1) modulation of finger pressure using a potentiometer, and (2) coordination of finger control with speech production. The current study provided a necessary, systematic investigation of finger pressure modulation using real-time visual feedback while using an EL with variable pitch control.

The first critical finding is an overall improvement in force target bar matching as indicated by 10.84% improvement in FMP from the first to the last session (Fig. 4). This finding is consistent with previous reports that concurrent, augmented visual feedback enhances motor learning of complex tasks. For example, Swinnen *et al.* (1997) found that when young adults (18–20 years) were provided with continuous, augmented visual feedback, the acquisition of a cyclical arm flexion and extension task was improved. Similarly, Wulf *et al.* (1998) have reported that participants aged 18–31 years performed better on a ski simulator task when provided with augmented visual feedback. More recently, Ossmy and Mukamel (2018) found that online, visually relevant feedback can enhance short term motor learning finger movement tasks. This suggests that visual feedback may be exploited for the training of a variety of tasks, including that related to the training use of the EL. Provided that manual dexterity problems do not exist for a speaker, our data suggest that visual feedback and auditory training may increase the likelihood that a speaker can successfully acquire frequency modulation skills with an EL to enhance their speech.

Despite the observed enhancement in performance, further research is warranted to better understand the effect of frequency change on speech performance. Previous studies have shown that for complex motor tasks, increasing the frequency of feedback and KR have resulted in enhanced performance (e.g., Guadagnoli and Lee, 2004; Wulf *et al.*, 1998). In the current study, the frequency of feedback did not vary; therefore, our data only speak to an enhancement in performance, not whether longer-term learning has occurred over this condensed training timeframe. As such, we would suggest that it is necessary for future studies to manipulate the frequency of feedback; for example, using a faded schedule to investigate the long-term effects of training. That is, feedback to the speaker specific to skill acquisition could be modified by either employing a variable or random feedback schedule or by systematically reducing feedback using a fixed feedback schedule. However, such adjustments in providing feedback would almost certainly

https://doi.org/10.1121/10.0000936

require that a relatively high level of skill acquisition was first demonstrated prior to eliminating trial-by-trial feedback. As well, studies of this type are needed to quantify the frequency of feedback (both visual and knowledge of results) in an effort to determine what levels are optimal for enhancing EL training, or can be optimized when paired with other EL signal adaptations (Liu *et al.*, 2006; Saikachi *et al.*, 2009; Tanaka *et al.*, 2014). We would also like to note that while the capacity to adjust the sensitivity of the Tru-Tone® on/off control does exist, the EL used in our study was maintained at the standard "pre-set" recommendation of the manufacturer. In practice, although adjusting the frequency of the EL is always recommended to optimize speech, modifications in the sensitivity of the on/off control is extremely uncommon. Thus, the sensitivity setting of our EL was consistent with that of all commercially available Tru-Tone® devices. Consequently, the use of this setting was seen to carry an added level of external validity to our findings.

The second finding of this study relates to the importance of the motor training task. When evaluating performance by task, the participant-speaker showed decreased performance during the SW and S tasks; of interest here is the fact that both of these tasks differed from the FB task since they required both matching force bars and simultaneously producing EL speech. A possible explanation for the decreased performance is that both tasks involve the simultaneous coordination of two activities: (1) pressing down on the force sensor with very fine control of pressure, and (2) producing EL speech. Additionally, both the SW and S tasks required that the EL device be coupled to the neck and held in the same position for the entire training session. This simultaneous coordination of multiple requirements increases the difficulty of the task compared to the FB task, which only involved the motor skill of matching force bars without generating EL speech. Further, the FB task did not require neck coupling because it did not involve the production of EL speech. This observation points to the possibility that coupling verbal pitch production and pressure control requires a new set of training tasks, all of which would be included in the speaker's ability to self-monitor changes via audition.

Based on the current data, it is unknown whether this enhancement in finger force control will also lead to improved production of intonation in real time (Saikachi *et al.*, 2009). That is, because intonation in verbal communication is the result of a complex set of coarticulatory events, there may have been a mismatch between coordinating finger pressure and the coarticulatory events necessary to produce intonation. One potential explanation for this mismatch is that the finger is not normally used in the production of speech. With this assumption in mind, other studies using EMG controlled EL devices on the neck that can activate muscles associated with speech have been reported (Goldstein *et al.*, 2007). Despite the closer connection to speech, the difficulty in coordinating speech production and muscle coordination (whether via the neck or finger) is evident. Therefore, controlling intonation during speech conversations requires faster and more accurate control than in laboratory speech tasks.

At present, it is unknown whether our protocol will lead listeners to perceive finger controlled EL speech as more natural sounding. However, it is unclear whether this difference might be a product of the control strategy or speaker specific characteristics. An alternative explanation is that the relationship between the signal (finger force or EMG) and F_0 has not been identified. In the current study, the relationship between finger force and pitch was found to be linear. This was based on static measurements and may not translate to the same relationship in real time speech production tasks.

The reduced performance on the SW and S tasks also may be explained by the chosen strategy of motor acquisition. That is, the S task was broken down into smaller "simpler" components: SW and FB tasks. This breakdown of complex skills is often referred to as the "part-whole transfer" strategy; this strategy depends on "part-practice," which is the division of a task into independent skills (Dubrowski *et al.*, 2005). Partial practice is in opposition to "whole practice," which is the learning of a task in its entirety (Dubrowski *et al.*, 2005). During the training protocol, the participant-speaker noted that he did not perceive a successive learning pattern across the three tasks (i.e., there was no identified build up from the FB task to the S task). Instead, he noted a benefit from repeating the FB, SW and S tasks over consecutive weeks of training. The observation that the participant-speaker performed best on the FB task further highlights the possibility that mastery of the FB task (a relatively simple skill) is not transferrable to more complex skills (S and SW tasks). This observation also supports of the previously described "whole-practice" theory of motor learning (Dubrowski *et al.*, 2005). However, the present findings are based on a single-subject. For this reason, no definitive conclusions can be drawn about which learning strategy would be most beneficial. We suggest that future studies on EL training paradigms consider replacing the small discrete tasks used in this study to affect frequency change with those that are more continuous and natural (i.e., running, sentence-based stimuli).

The type of training tasks may be influenced by individual motor requirements. In the laryngectomized population, restrictions in manual dexterity secondary to changes in upper extremity function (Boulougouris and Doyle, 2019) as well as changes to the compliance of neck tissue (Meltzner *et al.*, 2003) would be essential considerations. The potential influence of such factors on EL skill acquisition may necessitate the need to further partition specific training tasks, and potentially, expand the number of training trials in order to habituate one's skill in manipulating pitch control.

With respect to the design of this work, one of the limitations of the present study is the lack of full experimental control during the training process. In our attempt to design and structure our training tasks, we opted for a quasi-experimental design. Foremost to our decision-making was the fact that the sequential nature of the learning tasks that were undertaken as part of our EL training protocol did not permit a withdrawal phase. Thus, the acquisition of one skill and the inability to remove its influence from follow up tasks within the sequential training protocol and the associated process of learning is confounding. Due to this structural training concern, a descriptive assessment of the learning process was deemed to be most appropriate. Further, in designing the training protocol that we employed, it was our belief that a logical, systematic approach to training EL use in a manner that would ultimately be parsimonious with clinical instruction, while at the same time permitting detailed procedural observation, was necessary (Barlow *et al.*, 2009). Collectively, we believed that such an approach to identifying and monitoring EL skill acquisition (force and pitch matching) was warranted; the acquisition of such data could then inform future clinical training with those who are laryngectomized.

As a further consideration relative to the quasi-experimental design used, we were also aware that as structured, our training process did not allow for an experimental reversal. We believed that utilizing a reversal design was not possible. This decision emerged from of our inability to identify and potentially apply "experimental contingencies to an alternate behavior incompatible with the target behavior trained" (McReynolds and Kearns, 1983, p. 41). That is, we were unable to identify a similar, non-target task for which the EL training process could be applied. Consequently, a reversal design was not deemed to be feasible as part of our proof-of-concept assessment.

While potential threats to the internal validity of the current data were unlikely to be influenced by the common concerns of history or maturation, the sequential training process itself could be a confounding factor. However, the procedural and technical structure of our EL training paradigm demanded that our ability to better understand the process of skill acquisition and complex motoric control to identify potential flexibility in how future training programs are constructed was essential (Connell and Thompson, 1986). It is, however, important to acknowledge that we were also explicitly mindful in our effort to eliminate additional variables that could have a direct and potentially variable influence on the training process itself. This desire led us to initiate our project with a normal speaker who would not exhibit potential training restrictions that might be encountered with an individual who had a laryngectomy.

For example, and as previously noted, limitations in movement and control of the upper extremity and the hand/fingers (acts involving both gross and fine motor control), changes in neck impedance due to surgery and/or radiation, and so on, are common to those who have undergone total laryngectomy. This approach also was viewed to be of benefit as it sought to consider the potential influence of idiosyncratic behaviors or functional restrictions that might characterize a larger clinical population (Barlow *et al.*, 2009). In this regard, the potential influence of multiple sensory factors (visual, proprioceptive, haptic, and auditory) and their functional relationships on EL skill acquisition provide considerable clinical challenges. Consequently, we believed that some sacrifice in experimental control inherent to the training process as well as flexibility in the design would yield results that would guide the future clinical application of EL training of active pitch control. In this respect, we carefully considered the recommendations of "flexibility" in pursuing this proof-of-concept investigation in a manner that would allow us to better understand the present variables when transferred to applications with a clinical population (Connell and Thompson, 1986). Although the limitations outlined must be carefully considered and are of substantial importance to interpreting the present data, our findings provide a rich resource for future experimental explorations of EL training specific to volitional pitch control.

V. CONCLUSIONS

The findings of the current study demonstrate the importance of exploring different means of enhancing intonation control in EL speech. Although an improvement in force control was observed, our findings highlight the difficulty and complexity of training fine levels of force control using an EL. However, this is the first step in understanding the effectiveness of a standardized and systematic training protocol. As the goal of this work has served as a proof-of-concept, it does not address direct clinical outcomes. That is, force control and pitch variability using a pressure sensitive EL do not guarantee that a speaker will be able to create meaningful communicative contrasts in a real-life speaking situation. It is also important to stress that there is an important intersection between the ability for a mechanical device to effect variation, in this case, pitch control via an EL, and the user's capacity to not only acquire but master such a skill. This study further underlines the importance of implementing and testing a systematic approach to learning intonation when using an EL. While further research is required, the current study has provided the first step toward developing training protocol with the goal of enhancing EL speech.

ACKNOWLEDGMENTS

We would like to thank Xinya Lin for her help in the programming of experiment in MATLAB. We would also like to thank Griffin Laboratories for their donation of two pressure sensitive EL devices. Portions of this work were completed by the first author under the supervision of the other two authors while she was affiliated with Western University.

[1]mV is not a force measurement value, but rather, it is the converted version of the mechanical force achieved by the hardware setup.

Barlow, D. H., Nock, M. K., and Hersen, M. (**2009**). *Single Case Experimental Designs: Strategies for Studying Behavior Change* (Pearson Education, Boston), pp. 164–165.

Barney, H. L. (**1958**). "XLIV A discussion of some technical aspects of speech aids for postlaryngectomized patients," Ann. Otol. Rhinol. Laryngol. **67**(2), 558–570.

Bennett, S., and Weinberg, B. (**1973**). "Acceptability ratings of normal, esophageal, and artificial larynx speech," J. Speech Lang. Hear. Res. **16**(4), 608–615.
Boersma, P. (**2001**). "Praat, a system for doing phonetics by computer," Glot International **5**(9/10), 341–345, available at http://www.fon.hum.uva.nl/praat/.
Boulougouris, A., and Doyle, P. C. (**2019**). "Shoulder dysfunction and disability secondary to treatment for head and neck cancer," in *Clinical Care and Rehabilitation in Head and Neck Cancer* (Springer Nature, New York), pp. 397–411.
Connell, P. J., and Thompson, C. K. (**1986**). "Flexibility of single-subject experimental designs. Part III: Using flexibility to design or modify experiments," J. Speech Lang. Hear. Res. **51**(3), 214–225.
Doyle, P. C. (**1994**). *Foundations of Voice and Speech Rehabilitation Following Laryngeal Cancer* (Singular Publishing Group, San Diego).
Doyle, P. C., Shanks, J. C., and Keith, R. L. (Eds.). (**2005**). "Historical highlights: Laryngectomy rehabilitation," in *Contemporary Considerations in the Treatment and Rehabilitation of Head and Neck Cancer: Voice, Speech, and Swallowing*, edited by P. C. Doyle and R. L. Keith (Pro-Ed Publishers, Austin), pp. 17–58.
Dubrowski, A., Backstein, D., Abughaduma, R., Leidl, D., and Carnahan, H. (**2005**). "The influence of practice schedules in the learning of a complex bone-plating surgical task," Am. J. Surg. **190**(3), 359–363.
Fodor, J. A., and Garrett, M. (**1967**). "Some syntactic determinants of sentential complexity," Percept Psychophys **2**(7), 289–296.
Gandour, J., and Weinberg, B. (**1982**). "Perception of contrastive stress in alaryngeal speech," J. Phon. **10**(4), 347–359.
Gandour, J., and Weinberg, B. (**1984**). "Production of intonation and contrastive stress in electrolaryngeal speech," J. Speech Lang. Hear. Res. **27**(4), 605–612.
Gandour, J., Weinberg, B., and Rutkowski, D. (**1980**). "Influence of postvocalic consonants on vowel duration in esophageal speech," Lang. Speech **23**(2), 149–158.
Goldstein, E. A., Heaton, J. T., Stepp, C. E., and Hillman, R. E. (**2007**). "Training effects on speech production using a hands-free electromyographically controlled electrolarynx," J. Speech Lang. Hear. Res. **50**(2), 335–351.
Guadagnoli, M. A., and Lee, T. D. (**2004**). "Challenge point: A framework for conceptualizing the effects of various practice conditions in motor learning," J. Mot. Behav. **36**(2), 212–224.
Hillman, R. E., Walsh, M. J., Wolf, G. T., Fisher, S. G., and 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," Ann. Otol. Rhinol. Laryngol. **107**(5), 2.
Lee, M., Moseley, A., and Refshauge, K. (**1990**). "Effect of feedback on learning a vertebral joint mobilization skill," Phys. Ther. **70**(2), 97–102.
Lieberman, P. (**1967**). "Intonation, perception, and language," MIT Research Monograph.
Liu, H., and Ng, M. L. (**2007**). "Electrolarynx in voice rehabilitation," Auris Nasus Larynx **34**(3), 327–332.
Liu, H., Zhao, Q., Wan, M., and Wang, S. (**2006**). "Application of spectral subtraction method on enhancement of electrolarynx speech," J. Acoust. Soc. Am. **120**(1), 398–406.
Ma, K., Demirel, P., Espy-Wilson, C., and MacAuslan, J. (**1999**). "Improvement of electrolaryngeal speech by introducing normal excitation information," in *Proceedings from Sixth European Conference on Speech Communication and Technology*, Budapest, Hungary (September 5–9), pp. 323–326, available at https://www.isca-speech.org/archive/archive_papers/eurospeech_1999/e99_0323.pdf.
Magill, R. A., and Anderson, D. I. (**2007**). *Motor Learning and Control: Concepts and Applications* (McGraw-Hill, New York), Vol. 11, pp. 2–423.
MATLAB. (**2011**). Version 7.12.0 (R2011a) (The MathWorks, Inc., Nattick, MA).
McReynolds, L. V., and Kearns, K. P. (**1983**). *Single-Subject Experimental Designs in Communicative Disorders* (University Park Press, Baltimore), pp. 1–253.
Meltzner, G. S., and Hillman, R. E. (**2005**). "Impact of aberrant acoustic properties on the perception of sound quality in electrolarynx speech," J. Speech Lang. Hear. Res. **48**(4), 766–779.
Meltzner, G. S., Hillman, R. E., and Stevens, K. N. (**2001**). "Analysis of acoustic factors contributing to poor quality of electrolaryngeal speech," J. Acoust. Soc. Am. **110**(5), 2764.
Meltzner, G. S., Kobler, J. B., and Hillman, R. E. (**2003**). "Measuring the neck frequency response function of laryngectomy patients: Implications for the design of electrolarynx devices," J. Acoust. Soc. Am. **114**(2), 1035–1047.
Miller, S. E., Schlauch, R. S., and Watson, P. J. (**2010**). "The effects of fundamental frequency contour manipulations on speech intelligibility in background noise," J. Acoust. Soc. Am. **128**(1), 435–443.
Ossmy, O., and Mukamel, R. (**2018**). "Behavioral and neural effects of congruency of visual feedback during short-term motor learning," NeuroImage **172**, 864–873.
Rothman, H. B. (**1978**). "Analyzing artificial electronic larynx speech," in *The Artificial Larynx Handbook*, edited by S. J. Salmon and L. P. Goldstein (Grune, New York), pp. 87–111.
Saikachi, Y., Stevens, K. N., and Hillman, R. E. (**2009**). "Development and perceptual evaluation of amplitude-based F0 control in electrolarynx speech," J. Speech Lang. Hear. Res. **52**(5), 1360–1369.
Snodgrass, S. J., Rivett, D. A., Robertson, V. J., and Stojanovski, E. (**2010**). "Real-time feedback improves accuracy of manually applied forces during cervical spine mobilization," Man. Ther. **15**(1), 19–25.
Swinnen, S. P., Lee, T. D., Verschueren, S., Serrien, D. J., and Bogaerds, H. (**1997**). "Interlimb coordination: Learning and transfer under different feedback conditions," Hum. Movement. Sci. **16**(6), 749–785.
Tanaka, K., Toda, T., Neubig, G., Sakti, S., and Nakamura, S. (**2014**). "An inter-speaker evaluation through simulation of electrolarynx control based on statistical F 0 prediction," in *Proceedings of Signal and Information Processing Association Annual Summit and Conference (APSIPA), Asia-Pacific*, Chaing Mai, Thailand (December 9–12), pp. 1–4.
Therrien, A. S., and Balasubramaniam, R. (**2010**). "Timing and visual feedback constraints on repetitive finger force production," Exp. Brain. Res. **201**(4), 673–679.
Uemi, N., Ifukube, T., Takahashi, M., and Matsushima, J. I. (**1994**). "Design of a new electrolarynx having a pitch control function," in *Proceedings of 3rd IEEE International Workshop on Robot and Human Communication*, Nagoya, Japan (July 18–20), pp. 198–203.
Vaissière, J. (**2005**). "Perception of intonation," in *The Handbook of Speech Perception*, edited by D. B. Pisoni and R. E. Remez (Blackwell, Malden), pp. 236–263.
Watson, P. J., and Schlauch, R. S. (**2009**). "Fundamental frequency variation with an electrolarynx improves speech understanding: A case study," Am. J. Speech-Lang. Pat. **18**(2), 162–167.
Weiss, M. S., and Basili, A. C. (**1985**). "Electrolaryngeal speech produced by laryngectomized subjects: Perceptual characteristics," J. Speech Lang. Hear. Res. **28**(2), 294–300.
Weiss, M. S., Yeni-Komshian, G. H., and Heinz, J. M. (**1979**). "Acoustical and perceptual characteristics of speech produced with an electronic artificial larynx," J. Acoust. Soc. Am. **65**(5), 1298–1308.
Wulf, G., Shea, C. H., and Matschiner, S. (**1998**). "Frequent feedback enhances complex motor skill learning," J. Mot. Behav. **30**(2), 180–192.