

# Design and Evaluation of a Mechanically Driven Artificial Speech Device

**Tyler G. Tuttle**

Department of Mechanical Engineering,  
Michigan State University,  
428 South Shaw Ln,  
East Lansing, MI 48824  
e-mail: tuttle@msu.edu

**Byron D. Erath<sup>1</sup>**

Department of Mechanical and  
Aeronautical Engineering,  
Clarkson University,  
8 Clarkson Ave,  
Potsdam, NY 13699  
e-mail: berath@clarkson.edu

*This paper presents the design of a mechanically driven artificial speech device to be used by laryngectomees as an affordable alternative to an electrolarynx (EL). Design objectives were based on feedback from potential end users. The device implements a mainspring powered gear train that drives a striker. The striker impacts a suspended drum-like head, producing sound. When pressed against the neck, the head transmits sound into the oral cavity, allowing the user to produce intelligible speech. The dynamics of the vibrating head and sound pressure levels (SPLs) produced at the mouth were measured to compare performance between the device and a common, commercially available EL. The results showed comparable performance and sound output.*

[DOI: 10.1115/1.4038222]

## Introduction

Speech is a uniquely human attribute that allows for communication, creation of music, and expression of emotions [1]. Vocal communication is comprised of both voiced and unvoiced speech sounds. Human voiced speech, or phonation, is produced by fluid–structure–acoustic interactions [1], and is described by the aerodynamic–myoelastic theory of phonation. It postulates that vocal fold oscillations arise from the expulsion of air from the lungs, which produces aerodynamic forces that drive, and are coupled with, the elastic vocal folds [2]. These vibrations produce a pulsatile jet with a resulting pulsatile pressure field that acts as the sound source of voiced speech. The resonances of the vocal tract are manipulated through posturing of the oral cavity to modulate the sound source and form intelligible speech. The fundamental frequency of human voiced speech typically spans the range of approximately 80–220 Hertz [1]. Unvoiced speech sounds, such as fricatives and plosives, are produced as a result of airflow constrictions, such as in the pronunciation of /s/ in “miss” [3]. Voice disorders disrupt the normal speech process, and consequently, have debilitating impacts on vocal communication.

A laryngectomy is an invasive surgical procedure, whereby an individual loses the ability to phonate. A laryngectomy involves the removal of the larynx, usually as a treatment for laryngeal cancer [4], although acute trauma [5] may also necessitate the procedure. During a laryngectomy, the larynx is surgically excised, and the trachea is rerouted through a stoma that is created in the neck to enable breathing. Removal of the larynx has devastating consequences including: loss of speech; decreases in social and emotional engagement; loss of taste; and difficulty swallowing [6]. Laryngectomees have also been found to withdraw from conversations, feel stigmatism about the changes to their voice, experience embarrassment about their tracheostoma, and have increased anxiety and depression [7]. Overall, quality of life of laryngectomees has been shown to be lower than those who have undergone nonsurgical treatment for laryngeal cancer [4].

Following a laryngectomy, there are three options for speech remediation: Esophageal speech (ES), tracheoesophageal speech (TES), and the use of an artificial larynx (AL) [8]. ES is performed by ingesting air into the esophagus and expelling it in a controlled manner to induce vibrations of a constriction in the upper esophagus, known as the pharyngoesophageal segment (PES), which produces flow-induced self-oscillations that are

similar to the vibration of the vocal folds in normal phonation [9]. The pulsatile jet and resulting pulsatile pressure field produced by PES vibration are manipulated by the mouth and oral cavity to form intelligible speech [9]. ES is the least common form of speech remediation due to the difficulty in mastering this technique, which requires a considerable time investment, and prolonged training by a specialist [10]. Moreover, only approximately 50–60% of laryngectomees are able to develop esophageal speech after receiving appropriate speech training [9]. However, ES is the most common form of speech remediation in developing nations because other forms of speech remediation require the purchasing and maintenance of an ancillary prosthesis [8].

The implementation of TES involves a separate surgical procedure following the laryngectomy where a tracheoesophageal puncture (i.e., a hole) is created to connect the trachea to the esophagus. A tracheoesophageal prosthesis, or voice prosthesis, which acts as a one-way shunt valve allowing air flow to pass from the trachea into the esophagus, is inserted into the tracheoesophageal puncture [11]. When a subject occludes the stoma, air expelled from the lungs is forced through the voice prosthesis and into the esophagus. The resulting airflow incites self-sustained oscillations of the PES, similar to ES, thereby producing voiced speech [11]. The key difference between TES and ES is that in TES air flow from the lungs acts as the flow source, whereas in ES, an individual is required to continually ingest and expel air during speech production. While TES is the most common primary form of speech remediation in developed nations [8], it requires continual access to specialized care for regular prosthesis replacement. In addition, approximately one third of individuals are unable to successfully produce TES [12].

An AL is a hand-held external device that independently produces sound and transmits it into the oral cavity, allowing the user to produce intelligible speech through posturing of the oral cavity and lips. The earliest ALs were based on the concept of musical devices and incorporated a reed that was powered pneumatically by air from the tracheostoma to produce sound [11]. Pitch control was implemented by making the vibrating length of the reed adjustable [13]. These devices were mostly abandoned after the invention of the electrolarynx (EL), which is an AL that produces electromagnetically induced vibrations [11]. An EL contains a battery-operated voice coil that forces a striker to oscillate axially and strike a freely suspended head, thereby producing a harmonic sound source. When the EL head is pressed against the neck, the sound is transmitted through the soft tissue of the neck and into the oral cavity, thereby enabling speech. EL speech is the easiest form of artificial speech to learn [8] and is the most common speech remediation following laryngectomy [14]. Even though

<sup>1</sup>Corresponding author.

Manuscript received June 6, 2017; final manuscript received September 15, 2017; published online November 9, 2017. Assoc. Editor: Rita M. Patterson.

patients may eventually transition to an alternate form of voice production (ES or TES), ELs are typically the first remediation presented to patients postsurgery, and are often commonly employed as a backup device for TES users [12].

Consequently, a significant body of research has been dedicated to optimizing EL speech. This has included techniques such as improving the ergonomics of the operation through the implementation of hands-free operation [14–17], and addition of pitch control [18,19]. Efforts have also focused on optimizing the sound produced by EL devices, as they suffer from poor intelligibility due to un-naturally low fundamental frequency, and monotonic, or robotic, sound production [11,14,18,20,21]. To address these issues, the acoustic output from various ELs has been evaluated [21–24], and a number of strategies have been undertaken to improve sound production, including improving sound transmission through the neck by optimizing the sound output based on neck impedance [25,26], using adaptive noise canceling to reduce the effect of background noise associated with device use [27,28], implementing sound source control, and speech enhancement sub-systems [29], and incorporating automatic fundamental frequency control schemes [30–32].

While many of these approaches have been successful in enhancing acoustic output, a significant obstacle that prevents widespread global distribution of EL devices is the prohibitive cost, which can range from approximately \$300 to over \$1,000 (USD), with the additional cost of recurring battery replacement. This is particularly burdensome to achieving widespread deployment of ELs, especially in developing countries. The relatively high cost, as well as difficulty in addressing maintenance needs that may arise, can create a substantial financial burden for low-income individuals. Consequently, many individuals in developing countries are left without an option for speech remediation following laryngectomy [33].

An attractive alternative would be the development and implementation of a mechanically powered, low-cost AL (i.e., a mechanolarynx (ML)). While similar ideas have been proposed and pursued in the past [34],<sup>2</sup> they have been met with limited success, and there has been little to no quantification of their performance as an actual speech aid. The objective of this work is to develop and benchmark an ML that can be deployed as an affordable primary speech aid in developing nations, and as a backup device in developed nations.

## Design Objectives

Institutional Review Board approval was granted to survey individuals of the Utica, NY laryngectomy Support Group regarding the importance of features related to AL performance. Final design objectives were guided by the survey results as well as to ensure an acceptable level of device performance. The design objectives are specified in Table 1.

## Device Design

Figure 1 is an exploded assembly view of the ML design. The ML consists of a case (parts 1 and 2) that houses a mechanical actuator subassembly (part 14\*) that stores and transmits energy within the device. Figure 2 is the exploded view of the mechanical actuator subassembly. The mechanical actuator drives a striker (part 17, Fig. 2) that impacts a suspended head (parts 5 and 6, Fig. 1), thereby producing sound. The corresponding part descriptions and numbers are listed in Table 2. The function of each component of the ML design is discussed in detail in the following subsections (Mechanical Actuator, External Case, and Production and Assembly).

**Mechanical Actuator.** The source of power for the ML, contained within the mechanical actuator subassembly (Fig. 2), is a

<sup>2</sup><https://www.youtube.com/watch?v=5MAWNCA4LW0>

Table 1 ML design objectives

Design objectives	
Mass production cost per unit	<\$20
Usable run time	>10 s
Fundamental frequency	>60 Hz
One-handed operation	
No electrical components	
Pitch control	
High ease of use for elderly	

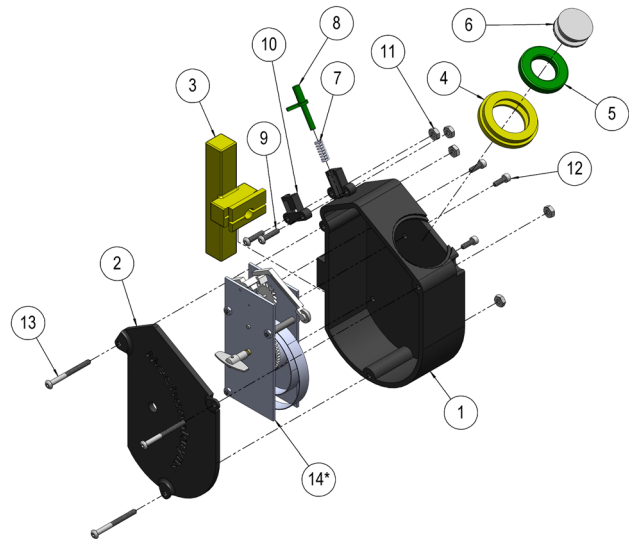


Fig. 1 Exploded assembly view of the ML, with corresponding parts numbered. See Table 2.

loop-end mainspring (part 16). The mainspring is wound manually by the user to store energy. It has a thickness of 0.43 mm, a width of 7 mm, a flattened length of 1.23 m, and is made of spring steel. An internal ratchet gear (part 20) allows for the mainspring to be wound without engaging the gear train. This increases the useful life of the gear train, allows for easier winding, and allows the device to be wound while in the “off” setting (i.e., without any motion of the subassembly gear train). Motion of the subassembly gear train, which transmits the energy from the mainspring to the striker (part 17), is controlled by a mechanical button that acts as an on/off switch (to be discussed in section case). When the

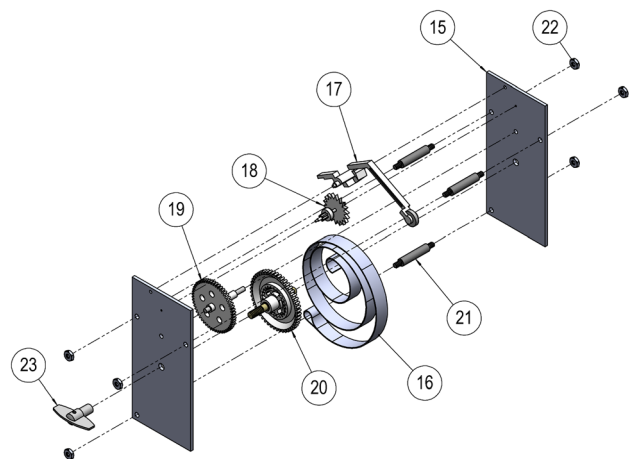


Fig. 2 Exploded assembly view of the mechanical actuator subassembly (part 14\* in Fig. 1). See Table 2.

**Table 2 List of the ML components with part numbers corresponding to Figs. 1 and 2 (part number with \* represents the mechanical actuator subassembly)**

Part #	Description	Quantity
1	Main case	1
2	Case front	1
3	Handle	1
4	Diaphragm housing	1
5	Diaphragm	1
6	Head	1
7	Compression spring	1
8	PCB	1
9	4–40 × 1/2 in rounded head screws	2
10	PCB cover	1
11	4–40 hexagonal nuts	5
12	4–40 × 5/16 in socket head screw	3
13	4–40 × 1–1/4 in rounded head screw	3
14*	Mechanical actuation subassembly	1
Part #	Part 14* Components	Quantity
15	Steel plate	2
16	Mainspring	1
17	Striker	1
18	Escapement wheel	1
19	Double gear	1
20	Internal ratchet gear	1
21	Male–male standoffs	3
22	Standoff nuts	6
23	Winding key	1

switch is toggled to allow rotation of the gear train, the spring energy is transmitted through the gear train (parts 18–20), to the escapement wheel (part 18), via a gear ratio of 54:1. The escapement wheel teeth interface with mating teeth is located on the striker (part 17). The striker consists of a nylon arm that has two teeth symmetrically placed about the point of rotation of the striker. An aluminum mass of approximately 0.25 gm is attached to the end of the arm of the striker. As the teeth of the escapement wheel engage with the teeth on either side of the striker, it causes rotation of the striker about the fulcrum point, raising the striker on the side of engagement. This, in turn, causes the opposing striker tooth that is symmetrically placed on the opposite side of the fulcrum to be brought down into contact with the escapement wheel. The continued rotation of the escapement wheel causes a see-saw motion of the striker about the fulcrum, causing the aluminum end of the striker arm to oscillate up and down, impacting the suspended head (part 6) and producing sound. The spacing of the teeth on the escapement wheel is such that the advancement of one tooth on the escapement wheel causes the striker to oscillate up and down, once. In this manner, a total of 18 oscillations occur per rotation of the escapement wheel. The gear transmission results in a striker impulse oscillation to mainspring rotation of 972:1. The gears, escapement wheel, and striker are manufactured from nylon. The mechanical actuation subassembly is held together by two steel plates (part 15) to enable accurate positioning of each component. The spacing between gear shafts is calculated as the sum of half of the pitch diameter of the interacting gears. Table 3 provides the specifications for each of the gears.

**Table 3 Gear specifications**

Part #	Gear type	Pitch diameter (mm)	Number of teeth	Tooth thickness (mm)
20	Internal ratchet gear	30	48	2.4
19	Large cog	26	54	2.0
19	Small cog	4.5	8	5.3
18	Escapement wheel	N/A	18	1.4
18	Gear	3	6	3.1
19	Striker	N/A	2	4.6

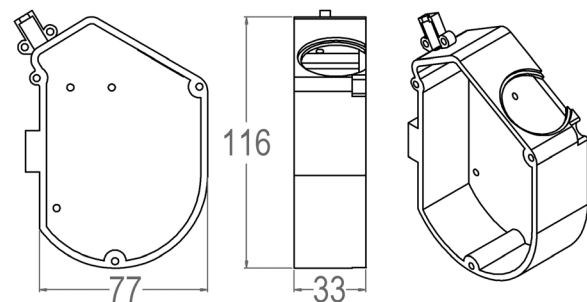
**External Case.** The mechanical actuator is contained within the case (parts 1 and 2). The case was designed to minimize the size, while allowing for full functionality of all features and components. The interior of the case contains enough open room for the fully unwound mainspring (part 16). The interior was lined with 0.318 cm thick Neoprene foam to reduce radiated noise from the escapement wheel–striker interface without greatly increasing the weight of the device. General dimensions of the case are shown in Fig. 3, demonstrating that the device is similar in size to other handheld objects, such as a coffee cup, and as such, is easily held in one hand.

A handle/wind assist (part 3) is included to hold the device and aid in winding the mainspring. The handle of the device can be slid into a slot on the main case (part 1) to hold the device during use, thereby facilitating one-handed use during speaking. To wind the mainspring (part 16), the handle is removed and the handle base interfaces with the winding key (part 23) that is attached to the internal ratchet gear (part 20). Prior studies have shown that elderly individuals are able to produce wrist torques of approximately 2.0–3.0 N·m [35]. The handle provides additional mechanical advantage so that users can easily produce the torque required (approximately 0.15–0.35 N·m) to wind the mainspring (part 16).

A pitch control button (PCB) (part 8) is also incorporated to allow modulation of the fundamental frequency of the ML. The PCB is located above the handle in a location that allows easy thumb activation. When the PCB is depressed by the thumb, the protrusion on the opposite end of the button contacts a tab that is rigidly attached to the striker. When the PCB contacts this tab, it constrains the range of motion of the striker (part 17) on the downswing, decreasing the time it takes for one oscillation, thus increasing the frequency of oscillation of the striker, and the resultant sound that is produced. When the PCB is fully depressed, the striker motion is constricted to the point that the range of motion of the striker teeth is smaller than the height of the escapement wheel (part 18) teeth, causing all internal device motion to halt. As such, the fully depressed position of the PCB acts as an “off” button/switch, whereby the user can pause usage to avoid wasting the spring energy, and/or if needed, wind the device to add more spring energy. The PCB has a tab that can be rotated to the side to fit into a groove on the PCB case (part 10) when the PCB is fully depressed, holding the PCB in the “off” position.

Device actuation begins by disengaging the tab while holding the PCB down with the thumb and then releasing the PCB. A small compression spring (part 7) disengages the button when no external force is applied (i.e., the thumb is lifted from the button). When the device is in operation, the user can release pressure on the button to achieve a lower pitch, or depress it to increase the pitch.

**Sound Production.** The ML produces sound in a manner similar to an EL. For the ML, the metal end of the oscillating striker contacts a head (part 6) that is suspended by a flexible diaphragm (part 5). The flexible diaphragm is an annular polyurethane rubber ring with an undeformed inner diameter of 15.88 mm (0.625 in),



**Fig. 3 Front, side, and isometric view of the main case (part 1), with basic dimensions in mm**



outer diameter of 28.58 mm (1.125 in), thickness of 4.76 mm (0.1875 in), and a durometer hardness of Shore 4000. The flexible diaphragm is positioned in a mirrored annular slot on both the housing (part 4) and the head such that it suspends the head. When contacted by the striker (part 17), this configuration amplifies the noise due to the compliance of the head, similar to the head of a drum, and also serves to isolate sound transmission from the main body (part 1) of the ML. As the striker oscillates radially, the metal mass on the end of the striker provides a contact impulse to the head, producing sound. During use, the vibrating head is pressed against the upper portion of the neck/lower jaw so that the generated sound is transmitted through the neck tissue and into the oral cavity, in the same manner as an EL.

**Production and Assembly.** All parts of the case (parts 1 and 2), vibrating head (parts 4 and 6), PCB (parts 8 and 10), and handle (part 3) were additively manufactured by way of material extrusion using a uPrint, model SE, with acrylonitrile butadiene styrene plastic. The steel plates (part 15) were manually machined from multi-purpose O1 tool steel. The mainspring (part 16), Neoprene foam that lined the case (not shown), flexible diaphragm material (part 5), gears (parts 19 and 20), escapement wheel (part 18), striker (part 17), compression spring (part 7), winding key (part 23), and fasteners (parts 9, 11–13, 21, and 22) were purchased.

Assembly of the ML is as follows: The Neoprene foam was cut to size and glued to the inside of the main case. The mainspring (part 16) is attached to the internal ratchet gear (part 20) by hooking a hole on the inner end of the mainspring to a notch on the internal ratchet gear shaft. The gears are attached to shafts with a shoulder on each end, with the smaller diameter shaft extension fitted into the gear shaft holes on the steel plates (part 15). The shoulders hold the gears in place by butting up against the steel plates, while the shaft/hole clearance was specified as a free running clearance fit to allow minimal rotary friction. The outer end of the mainspring is looped around one of the male–male stand-offs (part 21). The plates are bolted at a specified distance apart (17.15 mm) using male–male standoffs (part 21) and nuts (part 22). This assembly is then bolted to the main case (part 1). The PCB (part 8) and compression spring (part 7) are placed in position, and the PCB casing (part 10) is bolted to the main case. The front casing (part 2) is bolted to the main case. The flexible diaphragm (part 5) is fitted to the vibrating head (part 6), and inserted into the head casing (part 4), which is then attached to the main case with a cyanoacrylate adhesive. The winding key (part 23) is threaded to the internal ratchet gear shaft, and the handle (part 3) is attached to the case.

The mass of the assembled ML is approximately 210 gm, whereas an EL is approximately 125 gm. Although the mass of the ML is greater, it was deemed acceptable as it is within the range of common hand-held objects (e.g., an empty coffee mug has a mass of approximately 400 g).

## Methods

**Measurement of Vibrating Head Dynamics.** The dynamics of the ML head were measured and compared with that of a common EL (Blom-Singer electrolarynx digital speech aid) to assess the performance of the ML. The EL device can adjust both pitch and volume control; that is, the frequency and amplitude of the striker oscillations can both be independently controlled to optimize sound production. A laser distance sensor (LDS) (Wenglor PNBC 005) acquired the time history of the head displacement during operation for both devices. Data were acquired at a sampling rate of 4 kHz for the ML and 10 kHz for the EL, with the LDS providing a spatial resolution of 1.5  $\mu\text{m}$ . The sampling frequency was chosen from the preset values of the LDS to be approximately 100 times greater than the fundamental frequency of the ML and EL devices. Each device was secured vertically in a clamp with the head pointing upward; that is, in the orientation

in which it is commonly utilized. A small piece of electrical tape (negligible mass) was attached to the head, and a small amount of white correction ink was applied to the tape to improve the signal-to-noise ratio of the LDS. The LDS was positioned approximately 100 mm above the head of the device for each test, with the laser beam focused on the white correction ink on the device head. The LDS has a working range of 90–190 mm. The signal strength of the LDS, which varies based on the reflective nature of the measurement surface and the angle of incidence, was greater than 95% for all of the tests that were performed. Data were recorded for 90 s for the ML and 5 s for the EL. The longer time for the ML recording was to capture the decrease in fundamental frequency as the mainspring of the ML wound down. The highly periodic nature of the EL facilitated the shorter recording time. The ML was recorded at the highest obtainable frequency setting while the EL was recorded at two settings: (1) the lowest frequency and volume settings, and (2) the frequency and volume settings that optimize the intelligibility of the output, as subjectively defined by the user.

**Sound Pressure Level Measurements.** Sound output from both the ML and EL was quantified with a sound level meter (REED R8050), which provides an accuracy of  $\pm 1.4$  dB. Recordings were performed in the university music rooms, which utilize acoustic treatments on the walls to minimize acoustic wave reflection, providing a dry acoustic environment. The radiated sound pressure level (SPL) for each device and setting (ML—highest frequency, EL—lowest frequency and volume, and EL—optimal frequency and volume) was recorded for the following scenarios: (1) ambient background room noise, (2) device running with no neck contact so as to ascertain the maximum device noise produced, (3) device running and pressed against the neck with the mouth closed, thereby quantifying the background noise level of the device during operation, and (4) device running with the mouth open and postured to produce the vowel /a/, as in “father.” The vowel /a/ was chosen due to ease of repeatability. For each trial, SPLs were acquired over a period of approximately 10 s and both average and peak values were recorded. Prior to performing each measurement, the ambient SPL in the room was recorded to verify background noise levels were negligible. For all measurements, the sound level meter was mounted to a stand and placed 30 cm from the mouth of the user.

## Results and Discussion

The time history of the vibrating head displacement,  $\delta$ , is shown in Fig. 4 for a duration of 0.1 s for the cases of interest, namely: (a) the ML at the highest frequency setting; (b) the EL at the lowest frequency and volume settings, and (c) the EL at the frequency and volume settings that optimize the intelligibility of the output, as subjectively defined by the user.

The output from the ML reveals a highly periodic signal that is dominated by the impact of the striker contacting the vibrating head. Although clearly an underdamped system, the damping is sufficient enough that the time constant of the decay is much smaller than the period of the forcing function. Consequently, the dynamics of the vibrating head are well approximated as simply an impulse.

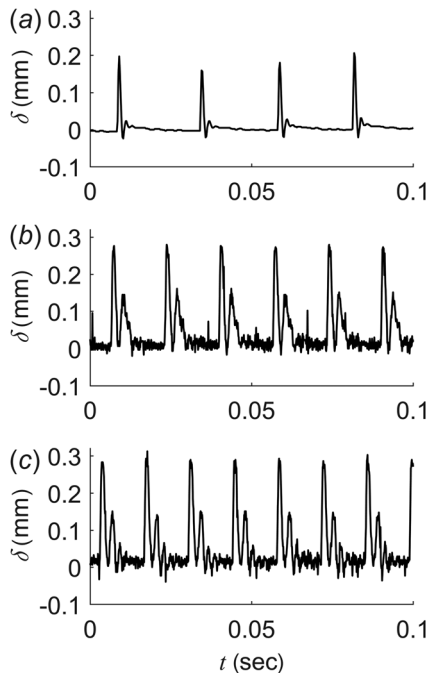
The head vibration of the EL exhibits higher harmonics, with an initial impulse corresponding to the impact of the striker, that is then followed by decaying oscillations, albeit with secondary peaks that are higher in amplitude than those of the ML. Note that the motion is not that of a purely second-order underdamped system, as the secondary oscillations are not symmetric about the rest position. This suggests that the initial contact from the striker produces a large impulse in the positive direction, but the oscillation of the head is then constrained as it moves in the opposite direction. This arises due to the design of the head, which, similar to the ML design, suspends a cylindrical head with a cylindrical ring of closed cell foam that is circumferentially seated in a groove

placed in the outer diameter of the head, and similarly seats in a groove along the inner diameter of the body. The inner diameter of the body below the groove is approximately 0.25 mm smaller than the inner diameter of the body above the groove. This has the effect of allowing minimally constrained oscillations in the positive direction after the striker impacts the head, whereby the amplitude of the head is only limited by the elastic energy of the foam ring. However, during rebound, once the head reaches the neutral rest position, the tighter clearance between the head and the body below the groove constricts the foam ring, increasing the effective stiffness. This rapid increase in the effective stiffness causes the motion of the vibrating head to be constrained in the negative direction, and to therefore rebound about this position. This has the effect of introducing oscillations with a non-zero mean, introducing higher harmonics into the signal.

The peak head displacement amplitudes were averaged over 120 oscillations, and are presented for all three cases of interest in Table 4. Data from the ML were computed from the first 120 cycles of oscillation to represent the highest amplitude output by the ML. Note that the mean of the ML peak amplitude (0.203 mm) is on order with the mean of the peak amplitudes of the EL at the lowest settings (0.265 mm) and at the optimal settings (0.299 mm). This is an important parameter to optimize as the amplitude of the head vibration can be directly related to the impact energy that is provided to the head, and therefore, the sound that is produced at impact that will be transmitted into the oral cavity for speech production.

Figure 5 shows the power spectral density (PSD) plots of the ML and EL for the three cases of interest. The first peak in each plot in Fig. 5 corresponds to the fundamental frequency of oscillation of the given case. The corresponding fundamental frequencies of oscillation for the three cases can be found in Table 4. The EL, with optimal settings, has the highest fundamental frequency (73.0 Hz), followed by the EL on the lowest settings (60.6 Hz), and finally the ML (39.7 Hz). The fundamental frequency of oscillation corresponds to the fundamental frequency of speech output that will be produced when using the device.

Modulation of the PCB of the ML was found to allow variation of the fundamental frequency over a range of approximately



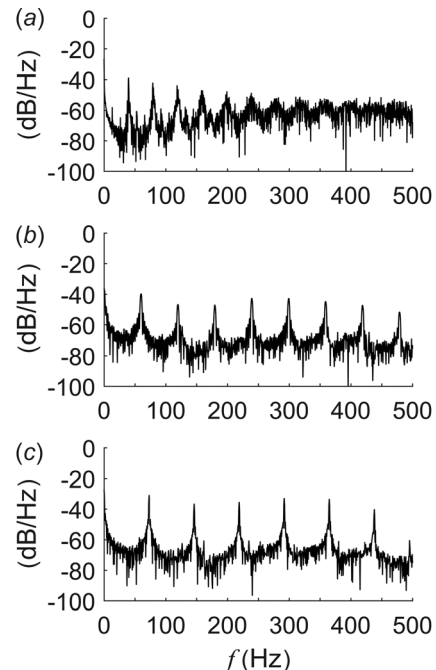
**Fig. 4** Time history of the head displacement for the first 0.1 s of runtime for the (a) ML, (b) electrolarynx on the lowest pitch and volume settings, and (c) electrolarynx on user-defined optimum settings

**Table 4** Fundamental frequency and mean peak amplitude of 120 oscillations for the ML, electrolarynx on the lowest pitch and volume settings, and the electrolarynx on the user-defined optimum settings

Artificial Larynx	$f_0$ (Hz)	Amplitude (mm)
ML	39.7	0.203
EL low	60.6	0.265
EL optimum	73.0	0.299

30–40 Hz. While the design goal of pitch modulation was achieved, it fell within an unnaturally low regime, and therefore, was not investigated further. In all cases, the fundamental frequency of the devices was unnaturally low when compared to physiological voice, which is typically approximately 80–220 Hz. For the ML, the achievable fundamental frequency was limited by the structural integrity of the gear train that transfers energy from the mainspring to the striker. Excessive torque resulted in skipping, broken gear teeth, and excessive wear. Nevertheless, as will be discussed later, while the unnaturally low frequency is undesirable for reasons of speech perception, it has a minimal impact on speech intelligibility. In addition, the achievable fundamental frequency is comparable with currently available commercial devices. Therefore, the reduced fundamental frequency was considered an acceptable compromise.

The PSD plot of the ML, shown in Fig. 5(a), has a greater amount of noise surrounding the peaks than in Figs. 5(b) and 5(c), which correspond to the EL. This is because the frequency of the ML head oscillation gradually decreases over time as the mainspring unwinds. The higher harmonics of the ML are markedly noisier and decrease monotonically, whereas the higher harmonics of the EL are more pronounced and do not decrease monotonically. As expected, the rebound of the EL introduces higher harmonics in the oscillations as was observed in the time-history plots of Fig. 4. This is important because the first two formants typically determine the sound of a vowel (e.g., /a/ as in “father” versus /Ov/ as in “code”) and the higher formants provide the timbre of the voice; that is, the personal aspect that makes individual



**Fig. 5** PSD of the head displacement for the (a) ML, (b) electrolarynx on the lowest pitch and volume settings, and (c) electrolarynx on the user-defined optimum settings

phonations sound different [36]. This is noteworthy as it suggests the EL will be capable of producing artificial voice that sounds closer to that of a human.

A spectrogram (Fig. 6) was computed from the time history of the ML head displacement to determine how the fundamental frequency of oscillation changes over time as the mainspring unwinds, and thus quantify the usable amount of speaking time available before the ML mainspring needs to be rewound. The spectrogram was computed using a window size of 1 s, with 50% overlap. While the total runtime of the ML following complete windup was determined to be approximately 300 s, considerable degradation of the signal is observed within the shorter 90 s time frame that is shown. During the first 90 s the fundamental frequency decreases from 40.1 Hz to 29.9 Hz, a difference of 10.2 Hz. It is observed that at approximately 40 s, there is considerable attenuation of the higher harmonics. Consequently, a run time of 40 s is deemed the limit of optimum usability without rewinding.

To quantify the acoustic output from the device, SPL measurements were acquired from the primary author of the manuscript for the three scenarios of interest. Although the author does not have a laryngectomy, sound transmission through the neck is comparable to that of a laryngectomee, and serves as a valid comparison for evaluating relative device performance. In all of the investigations the user took care not to phonate, but merely posture the mouth and oral cavity accordingly. During the SPL measurements, the ambient (background) SPL remained relatively constant between device trials, with a mean of 43.9 dB and a standard deviation of 0.18 dB. As previously discussed, because sound is radiated from the devices when not being used for speech, measurements were performed to quantify (1) the SPL radiated from the device when operated without pressing against the users neck, (2) the SPL radiated from the device when operated and pressed against the users neck (although maintaining the mouth closed), and (3) the SPL radiated from the device when pressed against the users neck, with the oral cavity postured to produce the vowel /a/, as in “father.” The results from the SPL measurements are presented in Table 5. The EL on optimum settings had the highest SPL output of 73.24 dB, and the highest SPL difference between vowel production and the device pressed against the neck with no speech (15.98 dB), which provides a simple measure of the signal to noise output. The ML performed second best for overall sound production (68.35 dB), but worst for the SPL difference between vowel production and the device pressed against the neck with no speech (6.31 dB). The EL on low settings performed the worst for overall sound production (60.10 dB), and second best for SPL difference between vowel production and the device pressed against the neck with no speech (11.06 dB).

These results indicate that the ML is capable of producing intelligible artificial voice, with SPLs that are comparable with those

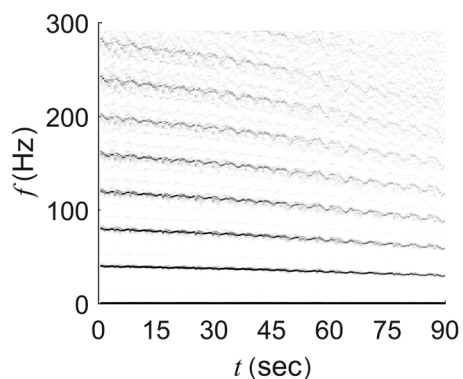


Fig. 6 Spectrogram of the ML head displacement for the first 90 s of runtime

**Table 5 Mean SPL measurements and (standard deviation) for the electrolarynx on low settings, the electrolarynx on the user-defined optimum settings, and the ML on the highest settings. SPL measurements were performed for the following cases: ambient background room noise (ambient); SPL of the device running with no neck contact (running); SPL of the device running and pressed against the neck with the mouth closed (neck); and SPL of the device running with the mouth open and postured to produce the vowel /a/.**

Artificial larynx	Ambient (dB)	Running (dB)	Neck (dB)	Vowel (dB)
EL low	44.02 (0.04)	61.68 (0.54)	49.04 (0.5)	60.1 (0.62)
EL optimum	43.86 (0.11)	73.64 (0.34)	57.26 (0.63)	73.24 (1.95)
ML	43.76 (0.23)	70.86 (0.58)	62.04 (0.8)	68.35 (1.33)

produced by commercially available devices. However, while overall sound production has been evaluated, speech intelligibility has only been cursorily addressed, as device testing and evaluation with end users cannot be performed without approval from the Federal Food and Drug Administration. Nevertheless, the ML results demonstrate the feasibility of such a device that would be capable of low-cost, mass market distribution to global regions that struggle with access to/affordability of commercially available artificial voice prostheses.

The cost to produce the final prototype was \$68.14, excluding labor. This cost can be greatly reduced by bulk purchasing, and exploring alternate manufacturing processes such as injection molding and die casting. Devices similar to the ML in size and mechanical complexity (e.g., wind-up alarm clocks) can be purchased for <\$10.00 (USD) at major retail stores, suggesting the objective of achieving a price point of <\$20.00 (USD) is feasible.

Shortcomings to be addressed in future endeavors include reducing the radiated background noise of the device and increasing the fundamental frequency of operation. The primary source of undesirable radiated background noise occurs at the interface between the escapement wheel and the striker where the escapement wheel teeth contact the teeth on the striker to drive the linear up-and-down motion. Various lubrication solutions were investigated, which were effective in significantly reducing the radiated noise from this interface. Unfortunately, in all cases, the improvement in noise reduction persisted for a relatively short amount of time, necessitating frequent reapplication of the lubrication. This was deemed an infeasible solution based on the objectives of a simple, low-cost, low-maintenance device. Future work will focus on redesigning the escapement wheel–striker interface to minimize the noise arising from impact. In addition, more robust gear design and housing tolerances would facilitate use of a higher torque mainspring, thereby allowing an increased fundamental frequency to be achieved without the added complications of increased tooth wear, and skipping. Finally, constraining the head motion so that it cannot symmetrically oscillate about its rest position would provide asymmetric forcing and should serve to excite higher harmonics, improving both sound transmission through the neck and amplification of the higher harmonics of the radiated SPL at the mouth, thereby increasing artificial speech intelligibility.

## Conclusion

An artificial speech device was developed that contains no electrical components. The device’s ability to produce intelligible speech illustrates that this approach has the potential to be implemented as a primary artificial speech device for individuals in developing nations, and as a primary or secondary speech device in developed nations. Vibrating head dynamics and SPL output of the ML show that it functions comparably to that of an existing, commonly used EL; that is, it produces similar dynamics at the vibrating head, and achieves similar acoustic output when utilized. Some performance characteristics can be improved to bridge the



performance gap between the ML and EL. By utilizing manufacturing techniques such as additive manufacturing, injection molding, and die casting, a device of this design has the potential to be manufactured at a more competitive price point than its electrical counterpart.

## Acknowledgment

The authors would like to thank the members of the Utica, New York Laryngectomy Support Group for their contributions.

## Funding Data

- National Science Foundation, Division of Chemical, Bioengineering, Environmental, and Transport Systems (Grant No. 1510367).

## Nomenclature

$f$  = frequency  
 $f_0$  = fundamental frequency  
 $t$  = time  
 $\delta$  = axial distance from equilibrium

## References

- [1] Mittal, R., Erath, B. D., and Plesniak, M. W., 2013, "Fluid Dynamics of Human Phonation and Speech," *Annu. Rev. Fluid Mech.*, **45**(1), pp. 437–467.
- [2] van den Berg, J., 1958, "Myoelastic-Aerodynamic Theory of Voice Production," *J. Speech, Lang. Hear. Res.*, **1**(3), pp. 227–244.
- [3] Maddieson, I., 2013, *The World Atlas of Language Structures Online*, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany.
- [4] Boscolo-Rizzo, P., Maronato, F., and Marchiori, C., 2008, "Long-Term Quality of Life After Total Laryngectomy and Postoperative Radiotherapy Versus Concurrent Chemoradiotherapy for Laryngeal Preservation," *Laryngoscope*, **118**(2), pp. 300–306.
- [5] Harrison, D. F., 1984, "Bullet Wounds of the Larynx and Trachea," *Arch. Otolaryngol.*, **110**(3), pp. 203–205.
- [6] Braz, D. S. A., Ribas, M. M., and Deditivis, R. A., 2005, "Quality of Life and Depression in Patients Undergoing Total and Partial Laryngectomy," *Clinics*, **60**(2), pp. 135–142.
- [7] Danker, H., Wollbrück, D., and Singer, S., 2010, "Social Withdrawal After Laryngectomy," *Eur. Arch. Otorhinolaryngol.*, **267**(4), pp. 593–600.
- [8] Xi, S., Li, Z., and Gui, C., 2009, "The Effectiveness of Voice Rehabilitation on Vocalization in Post-Laryngectomy Patients: A Systematic Review," *JBI Database System. Rev. Implementation Rep.*, **7**(23), pp. 1004–1035.
- [9] Van Weissenbruch, R., Kunnen, M., and Van Cauwenberge, P. B., 2000, "Cineradiography of the Pharyngo-esophageal Segment in Postlaryngectomy Patients," *Ann. Otol. Rhinol. Laryngol.*, **109**(3), pp. 311–319.
- [10] Quer, M., Burgués-Vila, J., and García-Crespillo, P., 1992, "Primary Tracheoesophageal Puncture vs Esophageal Speech," *Arch. Otolaryngol. Head Neck Surg.*, **118**(2), pp. 188–190.
- [11] Verkerke, G. J., and Thomson, S. L., 2014, "Sound-Producing Voice Prostheses: 150 Years of Research," *Annu. Rev. Biomed. Eng.*, **16**(1), pp. 215–245.
- [12] Chenausky, K., and MacAuslan, J., 2000, "Utilization of Microprocessors in Voice Quality Improvement: The Electrolarynx," *Curr. Opin. Otolaryngol. Head Neck Surg.*, **8**(3), pp. 138–142.
- [13] Riesz, R. R., 1930, "Description and Demonstration of an Artificial Larynx," *J. Acoust. Soc. Am.*, **1**(2A), pp. 273–279.
- [14] Heaton, J. T., Robertson, M., and Griffin, C., 2011, "Development of a Wireless Electromyographically Controlled Electrolarynx Voice Prosthesis," Annual

- International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Boston, MA, Aug. 30–Sept. 3, pp. 5352–5355.
- [15] Goldstein, E. A., Heaton, J. T., and Kobler, J. B., 2004, "Design and Implementation of a Hands-Free Electrolarynx Device Controlled by Neck Strap Muscle Electromyographic Activity," *IEEE Trans. Biomed. Eng.*, **51**(2), pp. 325–332.
- [16] Goldstein, E. A., Heaton, J. T., and Stepp, C. E., 2007, "Training Effects on Speech Production Using a Hands-Free Electromyographically Controlled Electrolarynx," *J. Speech Lang. Hear. Res.*, **50**(2), pp. 335–351.
- [17] Kubert, H. L., Stepp, C. E., and Zeitels, S. M., 2009, "Electromyographic Control of a Hands-Free Electrolarynx Using Neck Strap Muscles," *J. Commun. Disord.*, **42**(3), pp. 211–225.
- [18] Arifin, F., Sardjono, T. A., and Purnomo, M. H., 2014, "The Relationship Between Electromyography Signal of Neck Muscle and Human Voice Signal for Controlling Loudness of Electrolarynx," *Biomed. Eng. Appl. Basis Commun.*, **26**(5), p. 1450054.
- [19] Uemi, N., Ifukube, T., and Takahashi, M., 1994, "Design of a New Electrolarynx Having a Pitch Control Function," Third IEEE International Workshop on Robot and Human Communication, (RO-MAN), Nagoya, Japan, July 18–20, pp. 198–203.
- [20] Liu, H., and Ng, M. L., 2007, "Electrolarynx in Voice Rehabilitation," *Auris Nasus Larynx*, **34**(3), pp. 327–332.
- [21] Nagle, K. F., Eadie, T. L., and Wright, D. R., 2012, "Effect of Fundamental Frequency on Judgments of Electrolaryngeal Speech," *Am. J. Speech-Lang. Pathol.*, **21**(2), pp. 154–166.
- [22] Williams, S. E., and Watson, J. B., 1987, "Speaking Proficiency Variations According to Method of Alaryngeal Voicing," *Laryngoscope*, **97**(6), pp. 737–739.
- [23] 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), pp. 1298–1308.
- [24] Isshiki, N., and Tanabe, M., 1972, "Acoustic and Aerodynamic Study of a Superior Electrolarynx Speaker," *Folia Phoniatr. Logopaedica*, **24**(1), pp. 65–76.
- [25] Norton, R. L., and Bernstein, R. S., 1993, "Improved Laboratory Prototype Electrolarynx (LAPEL): Using Inverse Filtering of the Frequency Response Function of the Human Throat," *Ann. Biomed. Eng.*, **21**(2), pp. 163–174.
- [26] 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), pp. 1035–1047.
- [27] Liu, H., Zhao, Q., and Wan, M., 2006, "Application of Spectral Subtraction Method on Enhancement of Electrolarynx Speech," *J. Acoust. Soc. Am.*, **120**(1), pp. 398–406.
- [28] Li, S., Wan, M., and Wang, S., 2009, "Multi-Band Spectral Subtraction Method for Electrolarynx Speech Enhancement," *Algorithms*, **2**(1), pp. 550–564.
- [29] Houston, K. M., Hillman, R. E., and Kobler, J. B., 1999, "Development of Sound Source Components for a New Electrolarynx Speech Prosthesis," IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Phoenix, AZ, Mar. 15–19, pp. 2347–2350.
- [30] 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), pp. 1360–1369.
- [31] Tanaka, K., Toda, T., and Neubig, G., 2015, "An Enhanced Electrolarynx With Automatic Fundamental Frequency Control Based on Statistical Prediction," 17th International ACM SIGACCESS Conference on Computers & Accessibility (ASSETS), Lisbon, Portugal, Oct. 26–28, pp. 435–436.
- [32] Tanaka, K., Toda, T., and Neubig, G., 2014, "An Inter-Speaker Evaluation Through Simulation of Electrolarynx Control Based on Statistical F0 Prediction," Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA), Siem Reap, Cambodia, Dec. 9–12, pp. 1–4.
- [33] Tang, C. G., and Sinclair, C. F., 2015, "Voice Restoration After Total Laryngectomy," *Otolaryngol. Clin. North Am.*, **48**(4), pp. 687–702.
- [34] Creager, K. T., Goss, P. K., and Landauer, A. K., 2014, "An Inexpensive Mechanically Powered Laryngopharynx Excitation Device," Seventh World Congress of Biomechanics, Boston, MA, July 6–11.
- [35] Crawford, J. O., Wanibe, E., and Nayak, L., 2002, "The Interaction Between Lid Diameter, Height and Shape on Wrist Torque Exertion in Younger and Older Adults," *Ergonomics*, **45**(13), pp. 922–933.
- [36] Aronson, A. E., 1991, "Professional Voice: The Science and Art of Clinical Care," *Mayo Clin. Proc.*, **66**(12), pp. 1292–1293.