2010-06-23

Intelligibility of Electrolarynx Speech Using a Novel Actuator

Brian Madden
Dublin Institute of Technology, brian.madden@dit.ie

Mark Nolan
Dublin Institute of Technology, mark.nolan@dit.ie

Ted Burke
Dublin Institute of Technology, ted.burke@dit.ie

James Condron
Dublin Institute of Technology, james.condron@dit.ie

Eugene Coyle
Dublin Institute of Technology, Eugene.Coyle@dit.ie

Follow this and additional works at: http://arrow.dit.ie/teapotcon

Part of the Acoustics, Dynamics, and Controls Commons, Biomechanical Engineering Commons, Biomedical Commons, Biomedical Devices and Instrumentation Commons, Electrical and Electronics Commons, Otolaryngology Commons, Signal Processing Commons, Speech and Hearing Science Commons, and the Speech Pathology and Audiology Commons

Intelligibility of Electrolarynx Speech using a Novel Actuator

Brian Madden, Mark Nolan, Edward Burke, James Condron, Eugene Coyle

Department of Electrical Engineering Systems

Dublin Institute of Technology.

email : brian.madden@dit.ie

mark.nolan@dit.ie

ted.burke@dit.ie

james.condron@dit.ie

eugene.coyle@dit.ie

_____

During voiced speech, the larynx provides quasi-periodic acoustic excitation of the vocal tract. Following a laryngectomy, some people speak using an electrolarynx which replaces the excitatory function of the absent larynx. Drawbacks of conventional electrolarynx designs include the buzzing monotonic sound emitted, the need for a free hand to operate the device, and difficulty experienced by many laryngectomees in adapting to its use. Despite these shortcomings, it remains the preferred method of speech rehabilitation for a substantial minority of laryngectomees.

In most electrolarynxes, mechanical vibrations are produced by a linear electromechanical actuator, the armature of which percusses against a metal or plastic plate at a frequency within the range of glottal phonation. As part of the ongoing development of a hands-free alternative to the conventional electrolarynx, the authors have developed a novel actuator which is based on a lightweight pager motor similar to those used to produce vibration in many mobile phones.

In this paper, the intelligibility of speech produced using the novel actuator is compared to speech produced using a conventional electrolarynx. Three able-bodied speakers (two male, one female) performed a closed response test containing 48 monosyllabic words, once using a conventional electrolarynx and a second time using the novel actuator. The resulting audio recordings were randomized and replayed to four listeners who recorded each word that they heard. The results show that the speech produced using the novel actuator was substantially more intelligible to all listeners than that produced using the conventional electrolarynx.

The new actuator has properties (size, weight, shape, cost) which lends itself as a suitable candidate for possible hands-free operation. This is one of the research ideals for the group and this test methodology presented as a means of testing intelligibility. Any further actuators which exhibit desirable characteristics are also to be examined for electrolaryngeal intelligibility using this test procedure.

Keywords – laryngectomy, electro-larynx, intelligibility, pager motor.

_____

I   INTRODUCTION

A total laryngectomy is typically performed due to cancerous growths in the neck. Not only does it involve complete removal of the larynx, but also the trachea is disconnected from the pharynx and redirected through a permanent aperture in the front of the patient's neck (the *tracheostomy* or *stoma*), as shown in Figure 1. During voiced speech air is expelled from the lungs which provides the power source for excitation of the vocal tract, either through laryngeal phonation (voiced sounds), turbulence in a

vocal tract constriction (unvoiced sounds) or a mixture of both. In each case, the actual speech sound produced varies according to the configuration of the vocal and nasal tracts. Post total laryngectomy, normal speech is impossible because the conventional sources of vocal tract excitation are absent. The total laryngectomy procedure deprives the patient of their primary channel of communication. Since the loss of speech has an enormous impact on quality of life, speech rehabilitation is an important aspect of recovery following this surgery
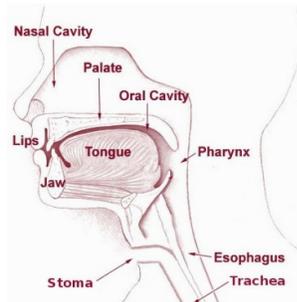

Figure 1: Redirection of airways following a TL [1].

a) Electrolarynx

Laryngectomees, as a part of their rehabilitation, are trained to communicate with as much ease as possible. Most are trained to use oesophageal or tracheo-oesophageal speech. For a minority, these channels of communication are not possible. For this remaining group, the most common form of communication is to use an external speech prosthesis. This is a mechanical larynx which uses an electromechanical actuator, i.e. the electro-larynx. The modern electro-larynx was invented by Harold Barney in the late 1950s [2]. It is a hand-held, battery-powered device which incorporates a transducer that generates mechanical pulses at a single frequency within the natural range of the human voice. The transducer uses an a coil-magnet arrangement that vibrates against a diaphragm when the output of an electrical oscillator is applied to its winding. The device is pressed against the mandible and this vibrates the pharynx which in turn resonates the air in the vocal and/or nasal tract. The vibrations are formed into speech by the articulators of the upper vocal tract.

Research to date has focused on the improvement of the quality of speech produced by the electro-larynx. Some significant contributions have been made by Houston et al. [3] who developed an electro-larynx which used digital signal processing to create a superior quality of sound. Shoureshi et al. [4] used neural-based signal processing and smart materials to improve the sound created. Liu et al. [5] and Cole et al. [6] focused on removing the buzzing sound created by the transducer. Uemi et al [7] in 1994 developed a

system that utilized measurements from air pressure that was obtained from a resistive component that was placed over the stoma to maintain the electrolarynx's fundamental frequency. Ma et al. in 1999 [8] used cepstral analysis of speech to replace the electrolarynx excitation signal with a normal speech excitation signal.

Despite all the acoustic improvements these studies have shown, they have however, been performed in isolation and have been deemed to be difficult to implement into the existing technology. Therefore, the basic concept of design which was first introduced by Barney et al. in the 1950's remains the same to this day. It has been shown that up to 50%-66% of all laryngectomees use some form of electrolarynx speech (Gray et al. [9], Hillman et al. [10]): either as a method of communication for speech rehabilitation post-surgery or as a reliable back-up in situations where esophageal or trcheo-esophageal speech is proving difficult.

b) Speech intelligibility

When determining the intelligibility of a speech signal, it is important to choose a suitable linguistic level at which to make measurements. Is it necessary to measure the accuracy at with which each phonetic element is communicated in order to assess whether each word is identifiable. It is also necessary to investigate whether the communication of a sentence is clear.

This type of linguistic dismissal can introduce an additional difficulty in that individual human listeners will ultimately differ in their capability to make use of these linguistic constraints. Even though it may be deemed necessary to assess the utility of a particular channel in order to convey the meanings of real spoken utterances, listeners will inevitably vary in their capacity to comprehend the speech, depending on their own linguistic ability. Many speech intelligibility tests consist of either phonetic unit, which are composed into: nonsense syllables, words which are used in isolation or in short sentences spoken in one breath for comfort Crystal et al [11] and Mitchell et al [12].

An issue that arises through the use of nonsense syllables is that many listeners could require training in order to be able to identify the component phonetic units, and they may be confused by phonemes which don't compare well with the spelling e.g. there, their, they're. Therefore by limiting listener reply's to real words thus allowing them to respond in ordinary spelling. This can however introduce other difficulties: firstly, that varying listeners may posses differing degrees of familiarity with the words that are being used; secondly, that some words are memorable and having heard a word once, some listeners may be biased in their usage of a particular word another time.

A possible solution to these problems includes the formulation of multiple word lists of reasonable

difficulty, allowing a listener to be used within a test more than once. Another option is to create tests consisting of closed response sets, making every listener needing to make the matching choices about the word which is under test.

Egan [13] pioneered one of the first lists of words for an intelligibility test in 1948. He created the list by using the concept of "phonetic balance" which meant that the relative frequency of the phonemes in the word lists corresponded to the relative frequency of phonemes in conversational speech. He constructed 20 lists containing 50 monosyllabic words and his intention was to balance average difficulty and range of difficulty throughout the lists whilst ensuring that the phonetic units that were present were represented equally.

II   METHODS
a) Novel actuator design

This design consists of a simple pager motor (which is typically found in a mobile phone) attached to a thin piece of high-density polyethylene (HDPE) by an aluminium support. When a current is sent through the motor, it causes the off-centred cam to rotate out of phase and causes an unbalanced centrifugal force. The motor in this design has a minute amount of play within the support and the HDPE section, thus causing a vibration that is resonated through the plastic. The thinner the plastic material is, the better the resonance becomes and resulting in a more efficient transfer of vibrations into the user's neck. Figures 2 and 3 shows the novel design for this concept and a region of a user's neck where it is envisioned that the final device will be attached.
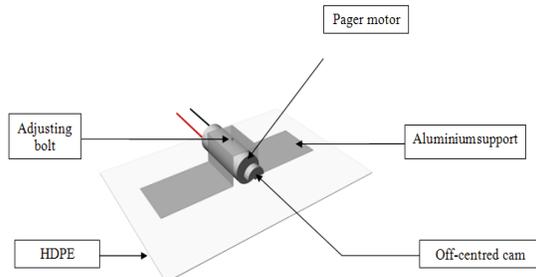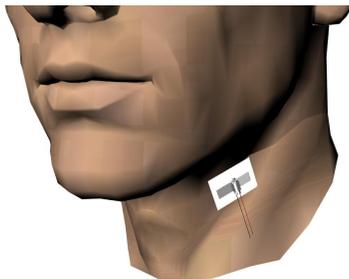

Figure 2: Novel motor design

Figure 3: Region where device is attached

b) Testing parameters

Three able-bodied speakers (i.e. non-laryngectomees), 2 male and 1 female, were chosen as participants and they received some basic pre-training in the use of an electrolarynx. They were instructed prior to recording to locate the point on their neck which produced the best resonance and thus the best sounding output (also known as the "sweet spot"). They were asked to hold their breath and maintain it held during each audio recording. Once while using the commercially available Servox electrolarynx and once while using the novel pager motor design.

A randomized sample of one of Egan's list of 50 words was taken and used for the intelligibility test bed. As there were only three participants, the 50 words were reduced to 48 so as to have a number of recorded samples of each device per speaker that was easily devisable by 6 i.e. 8 random words for each device and for each speaker.

The participants were instructed to sit upright in a chair and in order to keep the subject's posture constant during testing; their foreheads were supported in a head rest, figure 4. The height of the subject's seat was adjusted until an angle of 100 degrees from the chin to the torso was achieved. The microphone was then positioned 15cm away from the subject's mouth, figure 4.

Figure4: Experimental setup using head rest

c) Test Methodology

A combined microphone and preamplifier (Maplin KJ44X) was used to record the vocalization audio signals. The microphone was connected to a National Instruments 6023E 12-bit analog-to-digital converter (ADC). The ADC was set to a sampling frequency of 44.1 kHz on all channels.

Before testing, the pre-amplified microphone was calibrated using a Brüel & Kjær 2231 Sound Level Meter, at a distance of 15cm from a constant audio signal source. The audio intensity was adjusted and the output voltage from the microphone preamplifier was compared with the corresponding recorded Sound Pressure Level (SPL). A virtual instrument (VI) was created in LabVIEW which streamed in the data from the microphone through the ADC to the computer. The incoming data stream was broken into recordings of 5 second segments.

The VI gave a visual display of the recordings and they were saved as an .lvm file in a folder on the computer after recording of each utterance. Figure 5 illustrates a block diagram of the test set up.
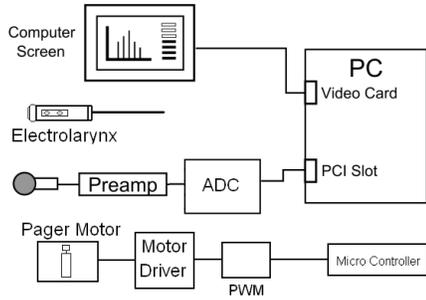

Figure 5: Block diagram of test set up.

d) Data Analysis

After the test was completed, each labview measurement file file was converted into a waveform audio file format file for convenience so as to be able to create randomized audio playlists containing the 48 recorded utterances and played to a listener on an audio player. Each waveform audio file format file was normalized to an audio level of -19dB on Cooledit Pro Version 5, with the out of band peaks selected as having no limits (i.e. not clipped).

Prior to creating the playlists, a formulated organisation of each one was arranged by taking a randomised selection of the 48 words for each individual playlist. Each word was then matched as shown in tables 1 and 2 below for example, if the first word on the playlist was "boil", the recording for "boil" was extracted from the recordings for each speaker and assigned to the word on the list.

Number	Speaker	Device
1 to 8	Speaker 1	EL
9 to 16	Speaker 1	Pager
17 to 24	Speaker2	EL
25 to 32	Speaker 2	Pager
33 to 40	Speaker 3	EL
41 to 48	Speaker 3	Pager

Table 1: Playlist order for listener 1

Number	Speaker	Device
1 to 8	Speaker 1	Pager
9 to 16	Speaker 1	EL
17 to 24	Speaker2	Pager
25 to 32	Speaker 2	EL
33 to 40	Speaker 3	Pager
41 to 48	Speaker 3	EL

Table 2: Playlist order for listener 2

Every second playlist was arranged so as to alternate the device being heard first, table 2. This was done so as not to create a listener bias towards the Servox electrolarynx as it was deemed that it could possibly take a number of recordings until the listener began to understand what to concentrate on.

III   RESULTS

The results for the each listener indicated a greater intelligibility towards the utterances which were spoken using the pager motor design. The results were tabulated subjectively using a 0 to 1 scoring system. This quantative analysis is presented in table 3 below. The overall intelligibility score across all three speakers for listener number one was twice that for the pager design, than the electrolarynx; pager – 82.3% intelligible and 62.5% intelligible for the Servox. Even when the utterances from the pager motor were in the first group heard by the listener number two, they were still greater than the next group of eight spoken using the Servox electrolarynx; pager – 64.5% intelligible and 36.5% intelligible for the Servox. Figure 6 illustrates the mean score obtained for both devices by each listener

Recorded utterance	Interpretation of utterance	Result	Score
Foil	Foil	√	1
Sail	Sale	✗	0.75
Baste	Paste	X	0.5
Pews	Huge	X	0.25
Whack	Wham	X	0

Table 3: Quantative analysis of point scoring system used.
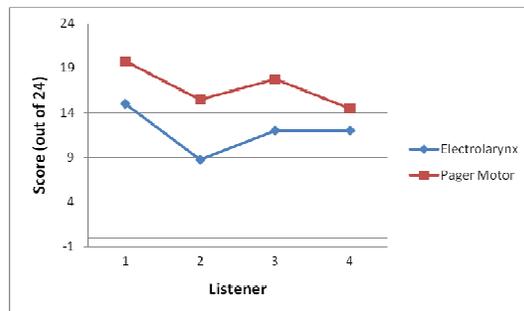

Figure 6: Mean scores of both devices for each listener

Figure 7 illustrates the percentage score in terms of the speaker's ability to use both devices. It shows that the pager motor performs better than the electrolarynx for every speaker. Even though there is a decline in ability for the pager motor, it still consistently remains easier to use than the electrolarynx.
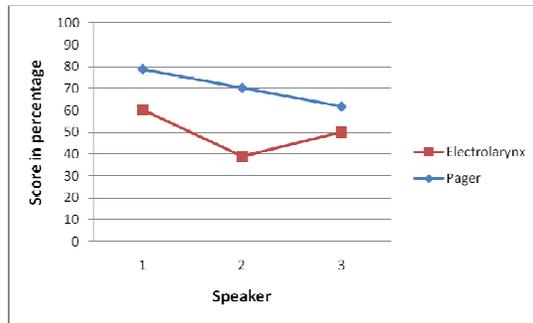
Figure 7: Plot of score in percentage for each speaker illustrating the ability to use each device

IV   CONCLUSION

This aim of this was to compare the intelligibility of speech that was produced using a novel actuator compared to that of speech produced using a conventional electrolarynx. The results illustrated that the speech produced using the novel actuator was substantially more intelligible to all the listeners than that produced using the conventional electrolarynx. There are many shortcomings with this study that are acknowledges by the authors such as; the few amount of users and volunteers who partook. However, this methodology is proposed as suitable test bed for a larger study in electrolaryngeal intelligibility. Initial results are encouraging and further work in the provision of a hands-free facility will be tested using this methodology.

REFERENCES

[1]http://training.seer.cancer.gov/ss_module06_head_neck/images/illu01_head_neck.jpg (as of 17/04/2010).

[2] Barney H.L, Madison N.J., 1963. *Unitary Artificial Larynx*. United States Patent Office. Patent No; 798,980.

[3] Houston K..M., Hillman R.E., Kobler J.B., Meltzner G.S., 1999. *Development of sound source components for new electrolarynx speech prosthesis*. IEEE Proc. International Conference on Acoustics, Speech, and Signal Processing Vol. 4, Issue, 15-19 Page(s):2347 - 2350 vol.4.

[4] Shoureshi R.A., Chaghajerdi A., Aasted C., Meyers A., 2003. *Neural-based prosthesis for enhanced voice intelligibility in laryngectomees*. IEEE/EMBS Proc. 1st Int. conference on neural engineering, Capri Island, Italy.

[5] Liu H., Zhao Q., Wan M., Wang S., 2006 *Enhancement of electrolarynx speech based on auditory masking*. IEEE transactions on Biomed. Eng. Vol. 53, No. 5.

[6] Cole D., Sridharan S., Mody M., Geva S., 1997 *Application of noise reduction techniques for alaryngeal speech enhancement*. IEEE TENCON – Speech and image technologies for computing and telecommunications. 491-494.

[7] Uemi, N., Ifukube, T., Takahashi, M., & Matsushima, J. (1994). Design of a new electrolarynx having a pitch control function, IEEE Workshop on Robot and Human, 198–202.

[8] Ma, K., Demirel, P., Espy-Wilson, C., & MacAuslan, J. (1999). Improvement of electrolarynx speech by introducing normal excitation information. Proceedings of the European Conference on Speech Communication and Technology (EUROSPEECH), Budapest 1999, 323–326.

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

[10] Hillman, R. E., Walsh, M. J., Wolf, G. T., Fisher, S. G., & Hong, W. K. (1998). Functional outcomes following treatment for advanced laryngeal cancer. Part I–Voice preservation in advanced laryngeal cancer. Part II–Laryngectomy rehabilitation: The state of the art in the VA System. Research Speech-Language Pathologists. Department of Veterans Affairs Laryngeal Cancer Study Group. Annals of Otology, Rhinology and Laryngology Supplement, 172, 1–27.

[11] Crystal, T. H., & House, A. S. (1982). Segmental durations in connected speech signals: Preliminary results. Journal of the Acoustical Society of America, 72, 705–717.

[12] Mitchell, H. L., Hoit, J. D., & Watson, P. J. (1996). Cognitive–linguistic demands and speech breathing. Journal of Speech and Hearing Research, 39, 93–104.

[13] Egan, J. (1948). Articulation testing methods. Laryngoscope, Vol. 58(9) , pp. 955-991