Dublin Institute of Technology
ARROW@DIT

Conference Papers

TeaPOT: People Oriented Technology

Spring 2011-01-26

Intelligibility of Electrolarynx Speech using a Novel Hands-Free 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 Biomedical Commons, Biomedical Devices and Instrumentation Commons, Computer Engineering Commons, Controls and Control Theory Commons, Electrical and Electronics Commons, Manufacturing Commons, Otolaryngology Commons, Signal Processing Commons, Speech Pathology and Audiology Commons, and the Systems and Communications Commons

Recommended Citation
Madden, B.:Intelligibility of Electrolarynx Speech using a Novel Hands-Free Actuator.International Conference on Bio-inspired Systems and Signal Processing,2011

This Article is brought to you for free and open access by the TeaPOT: People Oriented Technology at ARROW@DIT. It has been accepted for inclusion in Conference Papers by an authorized administrator of ARROW@DIT. For more information, please contact yvonne.desmond@dit.ie, arrow.admin@dit.ie.

This work is licensed under a Creative Commons Attribution-Noncommercial-Share Alike 3.0 License

INTELLIGIBILITY OF ELECTROLARYNX SPEECH USING A NOVEL HANDS-FREE ACTUATOR

Brian Madden, Mark Nolan, Edward Burke, James Condron, Eugene Coyle
Department of Electrical Engineering Systems, Dublin Institute of Technology, Dublin, Ireland
brian.madden@dit.ie, mark.nolan@dit.ie, ted.burke@dit.ie, james.condron@dit.ie, eugene.coyle@dit.ie

Keywords: laryngectomy, electro-larynx, intelligibility, pager motor, hands-free

Abstract: During voiced speech, the larynx provides quasi-periodic acoustic excitation of the vocal tract. 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 excitation. In this paper, the intelligibility of speech produced using a novel hands-free actuator is compared to speech produced using a conventional electrolarynx. Two able-bodied speakers (one male, one female) performed a closed response test containing 28 monosyllabic words, once using a conventional electrolarynx and a second time using the novel design. The resulting audio recordings were randomized and replayed to ten listeners who recorded each word that they heard. The results show that the speech produced using the hands-free actuator was substantially more intelligible to the majority of 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. This paper outlines the procedure for the possible testing of intelligibility of electrolarynx designs.

1 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 (National Cancer Institute, 2010). 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.

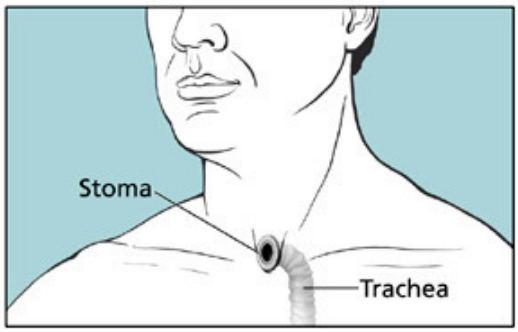

Figure 1: Redirection of airways following a TL

1.1 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 (Barney and Madison, 1963). 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. 1999) who developed an electro-larynx which used digital signal processing to create a superior quality of sound. (Shoureshi et al. 2003) used neural-based signal processing and smart materials to improve the sound created. (Liu et al. 2006) and (Cole et al. 1997) focused on removing the buzzing sound created by the transducer. (Uemi et al. 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. 1999) 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. 1976) and (Hillman et al. 1998): 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.

1.2 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 1982) and (Mitchell et al. 1996).

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 possess 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 et al. 1948) 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.

2 METHODS

2.1 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 causing 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 show the novel design for this concept and the region of a user's neck where it is envisioned that the final device will be attached.

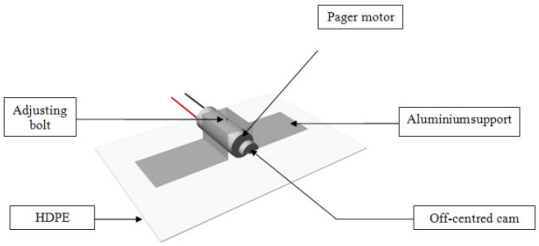

Figure 2: Novel motor design

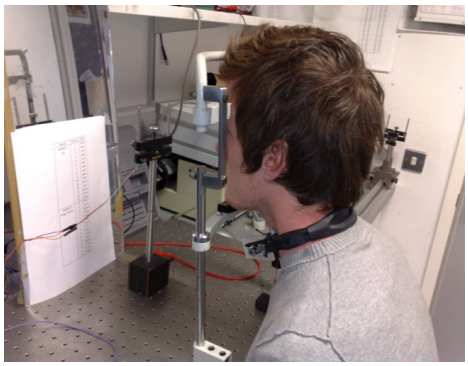

Figure 3: Region where device is attached

2.2 Testing Parameters

Two able-bodied speakers (i.e. non-laryngectomees), 1 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 30 words from one of Egan's list was taken and used for the intelligibility test bed. As there were only two participants, the 30 words were reduced to 28 so as to have a number of recorded samples of each device per speaker that was easily devisable by 2 i.e. 7 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.

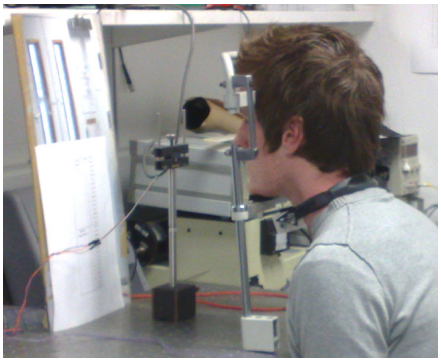

Figure 4: Experimental setup using head rest

2.3 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.

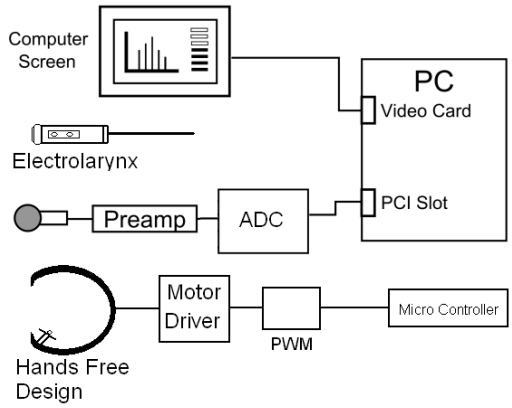

Figure 5: Block diagram of test set up.

2.4 Test Methodology

After the test was completed, each labview measurement file was converted into a waveform audio file format file for convenience so as to be able to create randomized audio playlists containing the 28 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 28 words for each individual playlist. Each word was then matched as shown in tables 1 and 2. For example, if the first word on the playlist was “leave”, the recording for “leave” was extracted from the recordings for each speaker and assigned to the word on the list.

Table 1: Playlist order for listener 1

Number	Speaker	Device
1 to 7	1	EL
8 to 14	1	Pager
15 to 21	2	El
22 to 28	2	Pager

Table 2: Playlist order for listener 2

Number	Speaker	Device
1 to 7	1	Pager
8 to 14	1	El
15 to 21	2	Pager
22 to 28	2	El

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.

2 RESULTS

The results for 90 percent of the listeners indicated a greater intelligibility of the utterances which were spoken using the hands-free pager motor design. The results were tabulated subjectively using a 0 to 1 scoring system. This quantative analysis is presented in table 3.

Table 3: Quantative analysis of point scoring system used.

Recorded utterance	Interpretation of utterance	Result	Score
Leave	Leave	√	1
Which	Witch	X	0.75
Towel	Dowel	X	0.5
Rude	Ruth	X	0.25
Exam	Sap	X	0

An average of the overall intelligibility score of all ten listeners for both speakers resulted in the novel hands-free pager design being 30% more intelligible that the commercially available Servox - 59% intelligibility scored for the pager and 30% intelligibility for the Servox. Even when the

utterances from the pager motor were in the first group heard by the listeners number 2, 4, 6, 8 and 10, they were still greater than the next group of seven spoken using the Servox electrolarynx; the pager motor design averaging at 78% intelligible and the Servox averaging at 55% intelligible. Figure 6 illustrates the mean score obtained for both devices by each listener, where listeners are represented on the x-axis and their cumulative scores are represented on the y-axis.

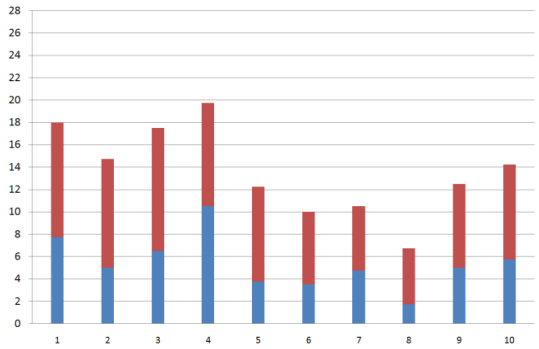

Figure 6: Mean scores of both devices for each listener. (Blue represents EL/Red represents Pager)

4 CONCLUSIONS

This aim of this was to compare the intelligibility of speech that was produced using a novel hands-free actuator compared to that of speech produced using a conventional electrolarynx. The results illustrated that the speech produced using the novel actuator were substantially more intelligible to all the listeners than that produced using the conventional electrolarynx. The limitations of this study are in the fact that only two speakers and ten listeners were used, a much larger variety of subjects (including a large number of electrolarynx users) would be required to create the random database of monosyllabic words for listeners to subjectively quantify. Further to this future intelligibility studies will not be limited only monosyllabic words and will be expanded to intelligibility of standardized phrases. Future paradigms of the future scoring system are to be explored as many listeners found it difficult to decipher certain words which have meanings with and without a vowel at the beginning e.g. “Bout” and “About” or “Wake” or “Awake”. Initial results are encouraging and further work in the provision of a second iteration of a hands-free facility which will be tested using this methodology.

REFERENCES

National Cancer Institute (2010). *What you need to know about cancer of the larynx.* Retrieved July 20, 2010, http:// www.cancer.gov/ cancertopics/ wyntk/ larynx/page2

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

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.

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.

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.

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.

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.

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.

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

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.

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.

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.

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