THE DISSERTATION ENTITLED

"How to make an Artificial Larynx by modifying Electro-larynx (Vocal Box)"

SUBMITTED TO

FLINDERS UNIVERSITY, ADELAIDE, AUSTRALIA

IN PARTIAL FULFILLMENT OF THE REQUIREMENT FOR THE

AWARDOF THE DEGREE

OF

MASTER OF ENGINEERING SCIENCE

IN BIOMEDICAL ENGINEERING

SUBMITTED BY

ASHINI NAILESHKUMAR RAMI

(2242007) (RAMI0036)

GUIDED BY

PROFESSOR DAVID POWERS

ASSISTANT PROFESSOR

Flinders UNIVERSITY ADELAIDE · SOUTH AUSTRALIA

DEPARTMENT OF ENGINEERING AND SCIENCE

FLINDERS UNIVERSITY, TONSLEY CAMPUS, ADELAID

AUSTRALIA

NOVEMBER- 2021

EXAMINER'S APPROVAL CERTIFICATE

ENGINEERING & SCIENCE DEPARTMENT

FLINDERS UNIVERSITY, ADELAIDE

This is to certify that the thesis entitled "How to make an Artificial Larynx by modifying Electro-larynx (Vocal Box)" submitted by Ashini Naileshkumar Rami (2242007) (rami0036) is approved for the award of the degree of Master of Engineering Science in Biomedical Engineering.

Signature of Examiners

External Examiner: ____________________

Date:

Place: Adelaide , Australia

ACKNOWLEDGEMENT

The project on "***How to make an Artificial Larynx by modifying Electro-larynx (Vocal Box)***" wouldn't have been a success without the guidance from various sources throughout the work.
First of all, I thank God almighty for His grace, blessings and the unforeseen guidance.

It gives me immense pleasure to express my heartiest gratitude to my thesis advisor, Professor David Powers for his unwavering support,mentorship and patience throughout this project. He was always available and very keen to help where needed with all his resources.

Secondly, I would like to thank Flinders university for giving me a chanceto do my thesis work in their esteemed laboratories, libraries around the devices which belongs to University.

Finally, I would thank my parents and all my friends for their valuable help and encouragement throughout the period of work.

ASHINI NAILESHKUMAR RAMI

ABSTRACT

During verbal communication, the larynx causes vibration due to its flexibility in the respiratory tract. Electrolarynxes or artificial larynxes are commonly used to restore speech after laryngectomy. An objective and subjective analysis of vowel quality is presented in this paper based on the design of a low-powered electrolarynx that uses a modified glottal source. A design is presented in the first part of the paper that can drive a variety of input signals. An effect of neck surface is pre-filtered to design a driving source in the second part. A comparison of power consumption between the prototype and a conventional electrolarynx is carried out. We compare the vowel qualities of the volunteers' vowels with those of normal vowels. According to our findings, the glottal modified wave helps reduce the amount of power required by the electrolarynx. Using the present approach, we measure the loudness, quality factor, and position of the formants as a measure of quality and find that they are better (or comparable) than traditional electrolarynxes. Electrolarynx with modified glottal waves are more power efficient than existing methods and have the potential to be incorporated into a wearable device. As well as providing better vowel quality than conventional driving signals, it also produces fewer noises.

Keywords: Electrolarynx, larynx, nltk, auditory, visual, voice restoration, esophageal speech, tracheoesophageal puncture, silent speech

Table of Contents

Table of Figure

CHAPTER 1: INTRODUCTION

1.1 Background

The total laryngectomy (TL) results in the loss of the patient's natural voice, which necessitates a major goal of speech rehabilitation. Generally speaking there are three types of methods to improve speech: esophageal speech (ES), tracheoesophageal speech (TES) and electrolarynx speech (ELS) [1. ES and TES both produce voice in the pharyngoesophageal (PE) segment of the throat, i.e., the source of voice is internal. A portion of the esophagus is administered air, which is then expelled. This results in mucosal vibration in the PE segment. TE fistulas or voice prosthetics are often used to create air channels. In contrast, esophageal content cannot reach the lungs through the esophagus.

The PE segment is driven by mucosal vibrations caused by pulmonary air in TES. An electrolarynx, a sound-producing device, mostly handheld, can be placed against one's neck or cheek, thus replacing the external sound source in ELS (Electrolarynx Sound System). The best speech rehabilitation method for restoring oral communication is not agreed upon worldwide based on science. According to some theories, TL patients who have better voice quality will also enjoy better quality of life [2, 3]. Multidimensional assessment is recommended to evaluate speech rehabilitation outcomes [4, 5]. Among the three substitute speech options, this systematic review compares acoustics, perception, and patient-reported outcomes (PROs). The pitch and amplitude of a voice are regularly measured in acoustic voice analysis [6]. Standard acoustic voice analysis does not always work when it comes to measuring substitute voices, which are characterized by having more noise components and less regularity than laryngeal voices [7].

In addition to the deviances in regularity compared to laryngeal voices, sensory evaluations of speech rehabilitation methods require a well-considered approach. The most suitable methods for evaluating substitute voices are to evaluate the quality of the voice and the intelligibility of the spoken word [8, 9]. The impact of speech rehabilitative treatments is typically evaluated with Quality of Life (QOL) questionnaires such as the EORTC QLQ-H&N35 and/or the EORTC QLQ-C30, which has questions about speech functioning [10, 11]. Speech rehabilitation results are better understood with PROs, such as the Voice Handicap Index (VHI) or Voice-Related Quality of Life (V-RQOL) [10–14].

A person's ability to speak naturally is terminated when the larynx is removed surgically as a result of laryngeal cancer [4]. Esophageal speech, tracheoesophageal speech, and electrolarynx are the three methods of re-establishing voicing without vocal chords and their space. By swallowing air, a person introduces air into the esophagus region and releases it abruptly into the oral cavity during esophageal speech [5]. Pharyngeal muscles vibrate as a result. The articulators convert these vibrations into speech. It requires a great deal of practice and training [6]. The one-way valve attached to a tracheal puncture is used in tracheo-esophageal speec [7]. By occluding the valve, the articulators direct air to the oral cavity, where it forms speech. In this method, there are hygienic problems such as fungal infections that can cause fluid leaks through tracheoesophageal puncture [8]. The electrolarynx is an electromechanical vibrator that replaces the larynx so that speech can be generated. During speaking, the device is held against the neck and a waveform generator generates a vibration. During speech production, the articulators move in response to vibrations of the device, converting vibrational energy into acoustic energy [9]. Electrolarynxes are also helpful for patients undergoing artificial ventilation when they are ill [10-12].

In addition to its conspicuous appearance, electrolarynxes have some other disadvantage The electrolarynx is a large device, which requires that the user hold it throughout verbal communication, causing inconveniences and awkwardness for the user. To make patients' lives easier, researchers are considering several miniaturization concepts. With the goal of reducing the size of the device, we designed a thin vibrator [13] that can be attached to the surface of the neck through the use of a brace. Using the wireless controller, the transducer can be controlled. The entire system is still heavy due to the requirement of 9 V supply. The wearable electrolarynx YOUR TONE II does not reveal the size or weight of its electrolarynx [14]. The motor of a tiny pager was used to implement a hands-free design [15]. In this application, the motor is attached to a thin membrane that pulses when voltage is applied. Pager motors have an insufficient handling capacity, so the vibration generated during speech is not audible. Speech intelligibility is affected by loudness reduction [16, 17]. Using a video camera and a tiny transducer, the hands-free design approach controls lip movement to enable electrolarynx control; however, it is not yet known whether the voice is audible [18]. Using mechanically driven gears, the artificial larynx has a fundamental frequency range of under 100 Hz [19]. This would result in a voice that is distracted from what is being spoken.

The current speech rehabilitation options have not been subject to a comprehensive review of the pros and cons. The collection of the best available evidence regarding the three speech rehabilitation methods would likely lead to a consensus as to which speech rehabilitation to use after TL and could assist clinicians, patients, and reimbursement agencies in making decisions. We investigated the acoustic, perceptual, and PRO effects of the three speech rehabilitation techniques following TL in this systematic review. In this research, we will investigate how the outcomes of various speech rehabilitation methods compare to those of normal laryngeal speech (healthy speakers), as well as what types of results are most favorable for each rehabilitation method. An examination of the literature on the outcome of speech following total laryngectomy (TL) was conducted using a systematic search strategy. In this search strategy, we focused specifically on the primary and secondary results that we were looking for. Depending on the literature, we selected the best primary and secondary results. The objective of the acoustic outcomes was to elucidate options for speech rehabilitation from objective data. Perceptual ratings and PROs served as vehicles for obtaining subjective information about the voices. In order to identify primary acoustic outcomes, we have selected fundamental frequency (Fo), harmonic to noise ratio (HNR), and voicedness percentage (%VO). Numerous authors have indicated that these outcomes are crucial to determining pitch, stability, and noise characteristics [7, 15, 16–17]. Other acoustic outcomes, such as jitter, shimmer, intensity, spectral tilt, and maximum phonation time (MPT), were interesting. The literature uses many of these outcome variables, although some are not as reliable in substitute voicing [16, 17].

The IINFVo scale was used to assess impression, intelligibility, noise, fluency, and voice quality, which are basic perceptual outcomes of interest. In addition to GRAAS, secondary perceptual outcomes of relevance were chosen from well-established perceptual assessment tools, such as unintended additive noise, fluency, and voicing functions [8, 18], and other recommended perceptual parameters of TL-speech in the literature. Among the most popular PROs are VHI13 and V-RQOL14. In addition, we included communication specific PROs on the EORTC QLQ-H&N35 [11] and the EORTC QLQ-C30 [10], which evaluate general quality of life including subsets related to communication.

1.2 Restoration of Voice History

Over 150 years ago, Czermak reported voice production in a patient with complete laryngeal stenosis when airflow was diverted through a reed tube from a tracheal cannula through the mouth. The patient's assistant created a custom speech apparatus consisting of a tracheostomy tube with a double lumen and an inlet extending into the pharyngostome for him, which was mounted onto a pneumatic device. In 1874, the German Company of Surgeons reported the successful result at its Third Congress in Berlin. Voice recovery advanced rapidly with esophageal speech reported in the mid-19th century, mechanical vibrations at the turn of the 19th century, and air conduits that enabled upper esophagus and pharynx to be reached in the mid-20th century, and tracheoesophageal puncture (TEP) speech that used bidirectional prosthetic valves in the mid-20th century [22][26].

1.3 Voice restoration in a nutshell

Voice occurs when lungs and larynx produce a sound or when the person speaks. Normal voice production depends on the following elements:

1. The lungs are responsible for generating air flows that flow through the larynx.
2. Apparatus for generating speech sound: the apposition of the paired vocal folds, in addition to air flow, creates vibrations in the vocal folds.
3. Phonetic voice is produced by modulating sound in the pharynx and oral cavity.

In a total laryngectomy (TL), the vibrating apparatus is removed, with the air generator and articulating tracts remaining. The airstream is diverted so that it does not pass through the articulating tracts [23][27]. The reason they are unable to produce sound is to do with this mechanism. A concomitant pharyngeal or tongue base disease may also influence the surgical excision of the articulating tract. By reintroducing a vibrating air column, which is then modified by the articulator, voice restoration is designed to artificially create a sound source [29][35]. The three approaches to reconstructing the vibrating apparatus and the air generator differ in the physiology of the alaryngeal voice, but the articulating tract differs for all three approaches. The vibrating apparatus used for electrolaryngeal voice production produces pharyngeal or oral cavity vibrations via an external electromechanical device. In large part, mucosal vibrations are not generated by air generators. Therefore, the mucosal vibrations that are caused by waves of air (and thus different from the tracheoesophageal or esophageal methods) are different. When using electrolaryngeal voicing, there is no air flow through the mouth during phonation, and aerodynamic studies indicate that the respiratory system is decoupled or less coupled to the voice during electrolaryngeal voicing [30].

In novice electrolarynx users, using exhalation during speech production mimicked their prechirurgical speech mechanics. Experienced electrolarynx users, however, decouple these functions and instead produce speech while holding their breath, improving speech acceptance. As there is no pulmonary requirement to make speech, there is no physiologic need to maintain exhalation during speech production. However, pulmonary air generators are modified differently for use with both the TEP and the esophageal methods [32]. Instead of TEPs, esophageal speech uses a one-way tracheostoma valve that allows pulmonary air to pass into the esophagus while oral air is delivered to and stored in the esophagus. The esophagus must be evacuated of air, at least partially, to trigger pharyngeal vibrations (neoglottis). There is a difference in voicing mechanics between these methods and electrolaryngeal voicing, as the produced mucosal wave is a direct response to an air stream. Many reports indicate that TL negatively impacts quality of life, especially psychosocial quality. Voice deprivation can also lead to social withdrawal and limit social relationships [37]. The reduction of sexual enjoyment and libido after laryngectomy or hypopharyngectomy is also common. In the aftermath of TL, restoring the voice quickly and effectively becomes a priority in order to prevent negative psychosocial and economic outcomes. There is evidence that voice restoration is possible for some patients who undergo laryngeal salvage with (chemotherapy) radiation; these patients demonstrate similar levels of quality of life to those who undergo TL without successful vocal rehabilitation. A preoperative speech therapy assessment is recommended for all patients planning to undergo TL [15][29]. Voice rehabilitation is facilitated by speech language pathologists, who help patients navigate the learning curves. To use an electrolarynx, speak esophageally, or use their TEP, patients need to learn how to do so. Financial considerations are also involved. Voice rehabilitation methods are very different in their costs, especially in third-world countries. Voice rating scales should be used both pre- and postoperatively to document long-term vocal dysfunction. Many people use the Voice Handicap Index and the University of Washington Quality of Life Scale. It is indispensable to have the ability to communicate with others to be able to go about one's daily activities and quality of life, and voice restoration after TL can provide this. However, there are few studies that have looked at electrolaryngeal speech solely from the perspective of quality of life, even though both microtechnology developments and upgraded user interfaces have significantly enhanced the quality of electrolaryngeal speech [11-17].

1.4 Definition of Electro-larynx

For a TL patient to regain voice, the electrolarynx was initially devised. It has an obvious advantage over synthesized speech and text-to-speech options as it allows the patient to maintain natural nuances of speech including the use of the oral cavity and preservation of articulatory abilities [12]. Consequently, there are a variety of electrolarynx devices, both traditional and modern. We list a few of those that are most commonly used.

1.5 Evaluation and History

In some academic settings, vibrating pneumatic devices were experimental nonelectric devices before the early 20th century. Although these instruments were impractical and unusable, their use was not widespread until the late 1920s, when the electrolarynx was developed. An electric powered device with a vibrating diaphragm was attached to a postlaryngectomy patient's neck. This device was the fifth installation and was later the prototype for many current electrolarynges with significant changes to overall size and portability [18][22]. The electric revolution began after World War II, and the Aurex Corporation developed the Neovox M-520 T, an electrolarynx that was smaller than the Western Electric 5A, but still required patients to remain stationary while using it. The transistor made the devices smaller, and Bell Laboratories created the first portable electrolarynx in 1959. This is still one of the first intraoral electrolarynges in commercial use today. It was developed in the 1980s by Cooper-Rand Electrolarynx (Luminaud Inc., Mentor, OH, USA). The handheld device weighs a pound [31].

1.6 Work Related

Only a few published studies have investigated the potential of EMG to detect speech. According to Chan et al. [1, 6], ASR was used to communicate with aircraft pilots on the myoelectric signal using an approach similar to the one proposed here. We embedded five bipolar electrodes in the oxygen masks of pilots and recorded the myoelectric signals generated during the acoustic pronunciation of the numbers "0" through "9". The utterances were also segmented using an acoustic signal. Based on a linear discriminant analysis (LDA) classifier, the authors reported a maximum word accuracy of 93% while using a hidden Markov model (HMM) classifier, the authors reported an accuracy of 86%. Additionally, they showed that the MES could enhance conventional speech recognition systems [1]. [2, 7] investigated how nonaudible speech could be recognized. By placing surface EMG electrodes below the jaw and on the larynx, they can intercept nerve signals that control speech muscles. Based on the MES and a Neural Network classifier, they demonstrated the ability to recognize non-audible isolated words [2]. Using six control words [2] and an extended vocabulary that included all ten English digits [7], they reported recognition rates of 92% and 73%, respectively. To accelerate the development of phoneme-based speech recognition, Jorgensen et al. extended their earlier isolated word experiments to recognize vowels and consonants. Additionally, they created a web-browser interface that is triggered by myoelectric signals [7]. For non-audible speech recognition, Manabe et al. suggested the use of rings-shaped electrodes wrapped around the thumb and two fingers. It is necessary to press the fingers against the face in a particular way for the electrodes to detect sEMG signals from facial muscles. In an analysis of conventional ASR techniques for recognizing the ten Japanese digits, the authors achieved a maximum recognition rate of 64% [9]. In the future, they hope that the system will develop into a mobile interface that can be used both in quiet and noisy environments.

As an alternative to relying on an acoustic signal, alternative methods are being investigated to overcome these limitations. The European Union funded this project as part of the IST programme. It was demonstrated by Chan et al. [1] that articulatory face muscles produce a myoelectric signal (MES) with sufficient information to enable discrimination of certain words accurately. These results hold even when the words are spoken inaudibly, i.e. without creating an audible signal [2]. To date, MES based speech recognition has been limited in its practicality. Firstly, the speaker's skin must be physically in contact with the electrodes.

Additionally, experiments are still limited to recognizing isolated words. As a final point, today's systems are very flawed, since they only work when training and testing conditions are the same. Similar to conventional speech recognition, the MES-based systems are heavily influenced by speech stylistics, voice rate, and pronunciation idiosyncrasies. Besides that, changes in electrode positions, temperature or even tissue properties can affect the myoelectric signal [3]. A session-dependent speech recognition system would be analogous to a conventional speech recognition system that is channel-dependent due to the speaker, microphone, and transmission of the acoustic signal.

MES-based speech recognition has been shown to result in significantly worse session-dependent performance loss than conventional systems do due to channel conditions. Despite this, MES-based speech recognition systems have only been developed for session dependent situations. The present paper will examine methods for adjusting data from a new recording session to prior training material, considering the session dependence. One of the biggest advantages of using the MES for speech recognition is that the words do not depend on the speaker pronouncing them audibly. Coleman et al. have demonstrated that controlling whispered speech and vocalizing speech is similar [4]. There are no studies that have examined the differences between audible and non-audible speech relevant to MES based speech recognition. The second focus of our research is to investigate these differences [17-21].

CHAPTER 2: Methods and Materials (Hardware)

2.1 OpenBCI

In late 2013, Joel Murphy and Conor Russomanno ran a successful Kickstarter campaign to fund OpenBCI, an open-source platform for brain-computer interfacing.

In addition to measuring and recording electrical activity produced by the brain (EEG), muscles (EMG), and heart (EKG), OpenBCI boards can be used with standard EEG electrodes. Alternatively, the OpenBCI boards can be used in conjunction with EEG signal processing tools open-source as well as the OpenBCI GUI [8].

TI's ADS1299 IC, developed for measuring biopotentials, is used in the 32bit OpenBCI.[2] Using an atmel ATmega328P IC (now deprecated), the 8bit board processes EEG data and writes it to a SD card, or sends it to software on a computer via Bluetooth [17].

Figure removed due to copyright restriction

Figure 1 Open BCI Hardware (Source “OpenBCI.com”)

A second Kickstarter campaign was launched in 2015 for the Ganglion board by OpenBCI. The device costs $200, has four input channels to measure EEG, EMG and EKG, as well as Bluetooth capability [28].

A Processing application written by OpenBCI for use with the OpenBCI has been made open-source. NodeJS and Python software for display and processing have also been released.

Tempt One, who has been diagnosed with ALS, has used the OpenBCI and the low-cost Eyewriter eye-tracking system to continue to draw using SSVEPs (Steady State Visually Evoked Potentials).

2.2 EEG

Electroencephalography (EEG) is a technique used to monitor the electrophysiology of the scalp, and has been shown to represent the activity of the brain's surface layer. Electrodes are typically placed along the scalp in a non-invasive manner. Invasive electrocorticography, or intracranial EEG, involves invasive electrodes [12].

Electrical activity of the brain is recorded with an EEG when multiple electrodes are placed on the scalp and voltage fluctuations result from ionic current. EEG is used for diagnostic purposes by focusing either on event-related potentials or on the spectral content of the EEG [39]. 'Stimulus onset' or 'button press' are examples of events which are time-locked. A frequency domain analysis of EEG signals analyzes the type of neural oscillations (popularly called "brain waves").

EEGs are commonly used to diagnose epilepsy, which can appear abnormal on an EEG.[2] They are also used for determining sleep disorders, depth of anesthesia, comas, encephalopathies, and brain death. As a first-line diagnostic tool for tumors, strokes, and other focal brain disorders, EEGs used to be widely used, but their use has declined with the advent of high-resolution imaging techniques like magnetic resonance imaging (MRI) and computed tomography (CT) [41]. EEGs continue to be valuable research and diagnostic tools, despite their limited spatial resolution. With millisecond temporal resolution, it is one of the few mobile technologies available. CT, PET, or MRI are all stationary techniques [42][44].

Among the derivatives of the EEG technique is evoked potentials (EP), which combine EEG activity with a stimulus of some kind (visual, somatosensory, or auditory). In cognitive science, cognitive psychology, and psychophysiological research, event-related potentials (ERPs) refer to averaged EEG responses that are time-locked to more complex processing of stimuli [38].

2.2.1 ERP Events

In more formal terms, an ERP is any stereotyped electrophysiological response to a stimulus that is a direct result of sensory, cognitive, or motor inputs.[1] By studying the brain in this way, one can evaluate its functioning without intervening in it.

Electroencephalography (EEG) is used to measure ERGs. Evoked potentials and induced potentials, representing subtypes of ERP, are magnetoencephalographic (MEG) equivalents of ERP [7].

Hans Berger discovered that one could detect the electrical activity of the human brain by placing electrodes on the scalp and amplifying the signal in 1924, the year the electroencephalogram was invented. A period of time can be plotted by plotting voltage changes. External stimuli can affect the voltages, according to his observations. Over the ensuing decades, the EEG proved an effective way to monitor brain activity [12][15].

Unfortunately, using pure EEG data made it difficult to isolate individual neural processes that are the focus of cognitive neuroscience. ERPs were more sophisticated ways of extracting sensory, cognitive, and motor events by averaging simple sensory, cognitive, and motor signals. They published their findings a few years later, in 1939, after recording the first ERP on awake humans in 1935-36. The 1940s were not known for much research on sensory issues because of World War II, but in the 1950s, research on sensory issues once again began to be conducted. A new era of ERP component discoveries began in 1964 when Grey Walter and colleagues described the contingent negative variation (CNV), a cognitive component. After Sutton, Braren, and Zubin (1965) described the P3 component, ERP component research became increasingly popular. As computers became more affordable in the 1980s, cognitive neuroscience research gained momentum. Currently, ERP is a widely used method in cognitive neuroscience research to study the physiological correlates of sensory, perceptual, and cognitive activities associated with processing information[5].

Figure removed due to copyright restriction

Figure 2 ERP Related Events (Source " http://faculty.washington.edu/losterho/erp_tutorial.htm")

Unlike ERP waveforms, which consist of positive and negative voltage deflections, ERP components are generally identified by a letter (N/P) indicating the polarity (negative/positive) and a number indicating their relative latency in milliseconds. An N100 is a negative-going peak that occurs at a latency of 100 ms (indicating that it is the first peak and is negative) and is often followed by a P200 or P2 (indicating that it is the second peak and is positive) [17]. Latencies are often quite variable for ERP components, especially those related to the cognitive processing. The peak time of the P300 component, for example, may be anywhere between 250 ms to 700 ms.

CRMs are frequently used in the fields of neuroscience, cognitive psychology, cognitive science, and psychophysiology. Numerous stimuli have been identified by experimental psychologists and neuroscientists as reliably eliciting ERPs from participants. According to some researchers, the timing of these responses provides insight into the timing of brain communication. When participants are exposed to the checkerboard paradigm shown above, the first response of their visual cortex is around 50–70 ms. This would seem to indicate that the brain decides when to receive a visual stimulus after the light enters the eye. In the oddball paradigm, for instance, the P300 response occurs at approximately 300ms regardless of the type of stimulus used: visual, tactile, auditory, olfactory, gustatory, etc. A general invariance is observed with respect to stimulus type, meaning that the P300 component reflects a higher cognitive response to unexpected or cognitively salient stimuli. Studies have also been conducted on the P300 response in the context of information and memory detection. [21] In addition, there has been research on P300 abnormalities in depression. P300 latency and amplitude are reduced in depressed patients[19].

It can be constructed a brain-computer interface which relies on the consistent response of the P300 to novel stimuli because of its consistency. As with the previous paradigm, if many signals are arranged in a grid, the rows are flashed at random, and then P300 responses are observed, the subject may be able to slowly type words by observing which stimulus he is looking at.

Research in the area of ERP is also being conducted in the area of efference copy. It is also important in human verbalization.[23][24] However, efference copies do not only occur during spoken expression, but are also present during silent speech production. This is also supported by event-related potentials.

ERPs such as the ELAN, N400, and P600/SPS are frequently used in research, especially in neurolinguistics. Additionally, machine learning algorithms are increasingly used to analyze ERP data[26][27].

2.2.2 SSVEP

Stable state visually evoked potentials (SSVEPs) are natural responses to visual stimulation at specific frequencies that appear in neurology and neuroscience research. An electrical activity is generated by the brain when stimulation of the retina ranges between 3.5 Hz and 75 Hz,[1].

With electroencephalography, this technique is widely used to study vision and attention. Research uses SSVEPs because of their high signal-to-noise ratio[2] and relatively low artifact sensitivity[3]. SSVEPs are also useful in identifying neocortical dynamic processes at the optimal frequencies. A stationary localized source and a distributed source are responsible for generating SSVEP.

Figure removed due to copyright restriction

Figure 3 SSVEP Graph Plotting (Source "https://www.researchgate.net/figure/An-SSVEP-BCI-system-with-frequency-encoding_fig2_264088415")

2.2.3 SSAEP

SSAEPs were recorded using depth and surface electrodes on rabbits to study steady-state auditory perceptions. When recording from the surface at a stimulus rate of 50 Hz, the SSAEP from the bregma was the largest and most representative. SSAEP surface potentials corresponding to the medial geniculate body were not present. [11] Furthermore, the latency of SSAEP in the inferior colliculus (IC) was quite similar to that of the surface potential. Further, for 50 Hz stimuli, the IC potential amplitude tended to become larger than for transient stimuli. Other auditory brain regions do respond to transient stimuli with amplitudes greater than the IC when they receive transients. The trapezoid body and auditory nerve do not amplify 50 Hz stimuli, however. This finding suggests that ICs play a significant role in the generation of SSAEPs [18].

Figure removed due to copyright restriction

Figure 4 SSAEP ERP Potential (Source https://www.hearingreview.com/hearing-products/accessories/components/auditory-steady-state-response-assr-a-beginners-guide**)**

2.3 EMG

Electrical muscle activity is measured and recorded with electromyography (EMG) through the use of an electromyograph. For each EMG, a record is produced called an electromyogram. During electrical or neurological stimulation, muscle cells generate electric potential, which is detected by an electromyograph. Analyzing the signals may reveal abnormalities, levels of activation, recruitment patterns, or the biomechanics of human and animal movements [12][24]. Electrodiagnostic medicine with needle electromyography is a technique popular among neurologists. In physiotherapy, kinesiology, and biomedical engineering, surface EMG is a non-medical technique used to assess muscle function. Computer Science also uses EMG as middleware in gesture recognition in order to allow human-computer interaction through physical actions.

Clinical and biomedical applications of EMG testing are numerous. Neuromuscular diseases are diagnosed with needle EMGs, or motor control disorders are studied with needle EMGs for research purposes. Injections of botulinum toxin or phenol into muscles are sometimes guided by EMG signals. Surface EMG is usually used for motion analysis and functional diagnosis. EMG signals are also used as control signals for prosthetic devices such as arms, hands and legs. Similarly, to get muscle signals from neck surface to analyse the movement of larynx during vowel generation, surface mount openBCI special dried electrode used. The electrode placement is generated as per the requirement which can be shown in the below figure.

Figure 5 Electrode Placement for EMG

As per the literature review and practical work, figure 5 illustrate the placement of an electrodes which provides good output data for the EMG signals during the speech generation. Although, some electrodes are place to measure the muscle activity near the eye during the speech generation as an observation if any signal can help to turn on device presicely.

CHAPTER 3: Methods and Materials (Software)

3.1 NLTK

Natural Language Toolkit, or NLTK as it is commonly known, is a library of programs and libraries designed for symbolic and statistical natural language processing (NLP) in English using the Python language. Stephen Bird and Edward Loper, both at the University of Pennsylvania, developed NLTK in 2008[4]. NLTK includes graphical demonstrations and samples. The toolkit comes with a book that presents the underlying concepts and explains the languages it supports,[5] as well as a recipe book.

Research and teaching need to be carried out in areas closely related to NLP, including cognitive science, artificial intelligence and information retrieval. [7] NLTK has been used successfully as a teaching tool, as a personal study tool and as a platform for prototyping new research systems. The NLTK is used in the courses of 32 US universities and 25 countries. Classification, tokenization, stemming, tagging, parsing, and semantic reasoning are all supported by NLTK.

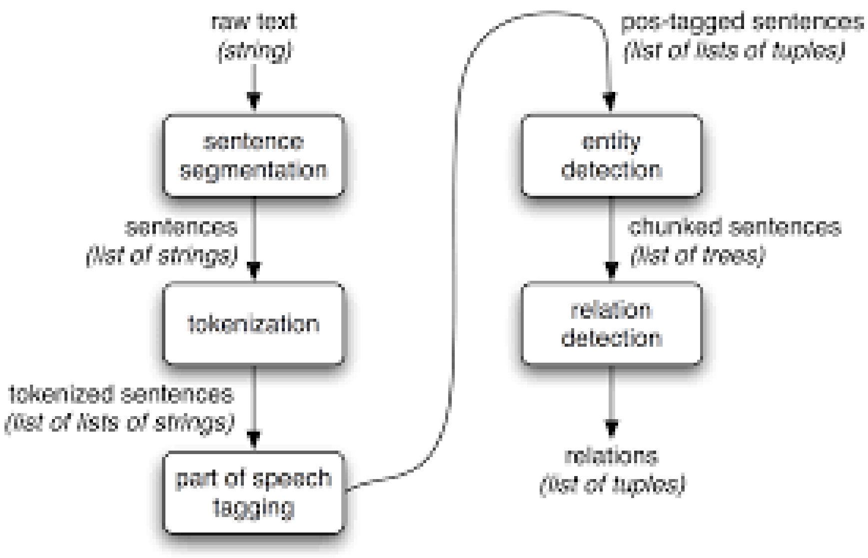

Figure 6 NLTK Working

3.2 Statistical Model in Python

Statisticians can conduct analyses to determine whether a dataset fits a certain distribution, in other words, to determine if the data corresponds to a particular theoretical model [17].

The analysis is known as distribution fitting and is based on the interpolation of mathematical functions that represent the observed phenomenon [21].

A potential example would be to have a set of observations x1,x2,xn... and you want to determine if those observations are indicative of a population described by f(x,*), where * is the vector of parameters to estimate based on the observations.

Statsmodels is the statistical library/module that Python programmers should know. With this library/module, you can perform multiple operations for statistical analysis using the SciPy Python library.

The api extension is a bit different than those found in most other Python libraries and modules. Statsmodels differs in many ways from other Python modules in terms of its nomenclature and syntax. X and Y are the variables in Statsmodels' endog and exog terminology when analysing data for statistical purposes [18][19]. A system's endogenous condition is implied by the term endog, which signifies a condition that is caused by factors within that system. The word exogeneous, on the other hand, basically means caused by factors external to a system. When using the Statsmodels module documentation, one should keep this terminology in mind.

How can Statsmodels be used in statistical modeling? Multivariate linear regression can be carried out using the Ordinary Least Squares(OLS) method using the OLS sub-module. WLSI is the submodule that needs to be used if Weighted Least Square(WLS) is performed. Those looking to perform should use the Generalised Least Square (GLS) sub-module. ANOVA tests, regressions with discrete dependent variables, linear mixed effects models, etc. are also possible in the context of regression using Statsmodels.

Based on the specific problem encountered, you can perform numerous statistical tests with the stats sub-module. There are a number of Chi-Square tests that can be performed with the stats sub-module, including Anderson-Darling, Ramsey's RESET test, and Omnibus tests for normality.

The Statsmodels module would not be complete without its graphics submodule. Also included in this module is the graphics submodule for plotting and visualizing statistical results.

Figure removed due to copyright restriction

Figure 7 An example of statistical modeling in Python (Source https://towardsdatascience.com/statistical-modelling-with-python-the-three-must-know-s-modules-79fa393e5640**)**

3.3 Neural Network

The term "neural network" generally refers to an artificial neural network, or a network of neurons. This is both a biological neural network, composed of biological neurons, or an artificial neural network used to devise artificial intelligence (AI) solutions [19][29]. The connections between neurons in a biological system are modeled as weights between nodes in artificial neural networks. Negative weights indicate inhibitory connections, while positive values reflect excitatory connections. The inputs are then summarised according to their weights [8]. Linear combinations are therefore performed. Activation functions control the output amplitude. As an example, an acceptable output range is generally between 0 and 1, but it could also be between 1 and *1.

Artificial networks may be used for predictive modeling, adaptive control, and applications where a dataset can be used to train them. Networks, which can draw conclusions from seemingly unrelated information set, can learn from their experiences through self-learning.

A biological neural network is constructed by chemically connecting or functionally associating neurons. In a network of neurons and connections, each neuron may be connected to many others. It is known that synapses are usually formed when axons connect to dendritic fibres, although other connections such as dendrodendritic synapses are possible. The diffusion of neurotransmitters results in other types of signaling, in addition to electrical signaling [17].

Biological neural systems process information in a similar manner to how artificial intelligence, cognitive modeling, and neural networks process information. Intelligent computing attempts to approximate some aspects of biological neural networks [44]. Artificial neural networks are used in artificial intelligence to build software agents (in computer and video games) or autonomous robots. They have proven useful in speech recognition, image analysis, and adaptive control.

Von Neumann's model gave rise to digital computers, which operate through explicit instructions executed via access to memory by a number of processors [26]. In contrast, neural networks have their origins in the attempt to model information processing in biological systems. The von Neumann model separates memory from processing, but neural network computing does not.

Both the theory of neural networks and artificial intelligence have served to improve our understanding of how neurons function in the brain.

In artificial neurons, neural networks are called artificial neural networks (ANN) or simulated neural networks (SNNs). Neural networks are interconnected groups of natural or artificial neurons, connected by mathematical or computational models that utilize a connectionistic approach to computation for information processing. As a rule, ANNs are adaptive systems that change their structure in response to external and internal signals [22-26].

Figure removed due to copyright restriction

Figure 8 Neural Network (https://wikipedia.com**)**

Statistical data modeling and decision making typically use neural networks, which are non-linear statistical tools. Modelling complex relationships between inputs and outputs or finding patterns in data can be accomplished using them [17].

In artificial neural networks, a set of simple processing elements (artificial neurons) are connected, which results in complex behavior based on the relationships between these elements. Logician Walter Pitts and neurophysiologist Warren McCulloch proposed artificial neurons for the first time in 1943, at the University of Chicago.

Recurrent Hopfield networks are a classical form of artificial neural networks.

Interestingly, Alan Turing first wrote about neural networks in his 1948 paper Intelligent Machinery, in which he described them as "B-type unorganised machines".

AINN models are useful because they can be used both to infer and use functions based on observations. Alternatively, unsupervised neural networks can be used to learn representations of inputs that capture points of the distribution, such as Boltzmann machines (1983), and recent deep learning algorithms that learn the distribution function of observed data implicitly. Using neural networks to learn is especially useful when the complexity of the data or task makes hand-designing such functions inappropriate.

CHAPTER 4: Results & Discussion

With the help of OpenBCI data we obtained the following results for the EEG data processed. We used OpenBCI sample data to induce ERP related potential events. Using OpenBCI cyton board, EMG data processed while speaking to fing the muscle signal activity during the vowel generation.

4.1 Results

4.1.1 EEG – ERP data output explanation:

Every classifier training was based on the same number of exemplars, namely thirty, from the same experiment in order to ensure comparability. The round robin procedure was applied each time training and testing were done on the same session. Training data was divided into a disjoint set of training sets each containing thirty exemplars of each vocabulary word (when the testing session differed from the training session), and the results were summed for the training sets.

The best results were achieved by speaker S2. Several non-audible sessions were already recorded before this participant came to the study. Through his years of experience, he developed a specific style of non-audible speak. Interestingly, we found that all speakers performed better as their level of experience increased. Compared to the results in Table 1, individual channels show a significant difference in performance. For all speakers, the best results can be obtained from channels EMG1 and EMG3. The two channels are corresponding to two distinct muscle groups, presenting orthogonal information. At the 9.56E-01 * 100 % level, we find an extremely significant performance improvement between two and three electrodes, while the performance difference between five, six, and seven electrodes is negligible.

Servox Digital Speech Aid uses a vibrating head to act as a transducer. The duty cycle of the driving signal can be adjusted based on the proportional tone. The vibrator is placed on the neck and the articulators are fixed so that vowels can be produced. A glottal modified wave and square wave are used to produce sustained vowels. Speech derived from these two drives is captured and digitized with a speech recorder. Initially, the prototype is measured to determine how much power it consumes. As a function of time, each device draws its own current and drops its supply voltage. Calculating the power consumption of the circuit is based on these readings. A second experiment is performed using a square wave as the excitation source. Vowel characteristics like formants, bandwidths, amplitudes, and spectral regions are plotted and compared to the normal vowel plot.

This dataset has realistic digitized 3D sensor locations saved as part of the .fif file, so we can view the sensor locations in 2D or 3D using the plot_sensors method.

A projector has already been added to the EEG data as an EEG common average. Plotting raw data with and without the projector lets us observe the effect on the raw data.

We can click on the graph to drop epoches or we can automatically do it. We have automatically done it afterwards.

Using the barplot, we determine which channels contributed most to the rejection of epochs. A channel that consistently leads to epoch rejections may be worth marking as bad in the Raw object and then rerunning epoching.

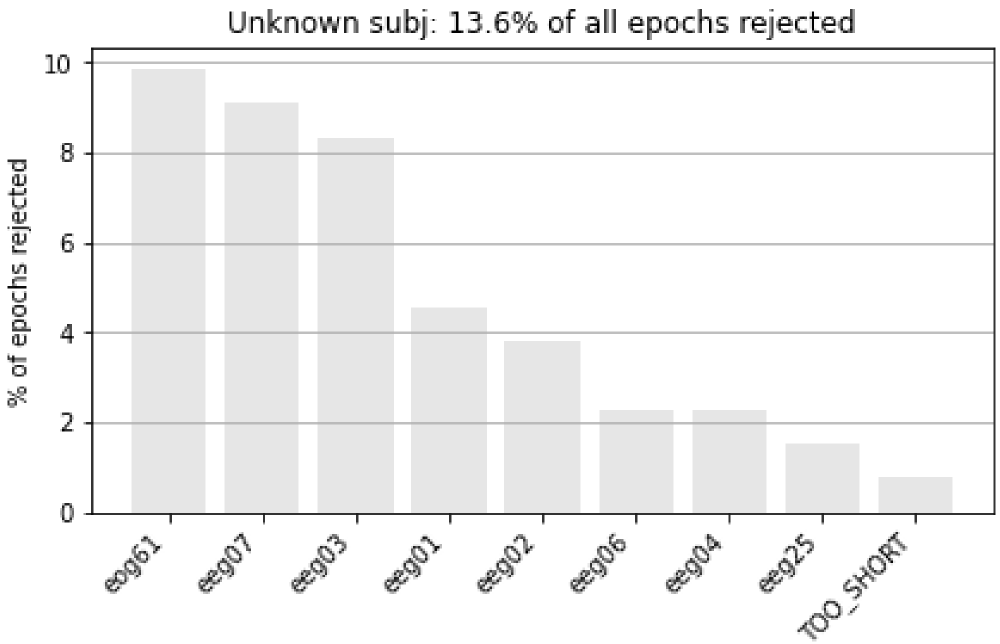

Figure 9 Rejected Epoches Bar Plot

Now we took evoked epoches and found the average to be used to plot a graph. This is done for both audio and video plot.

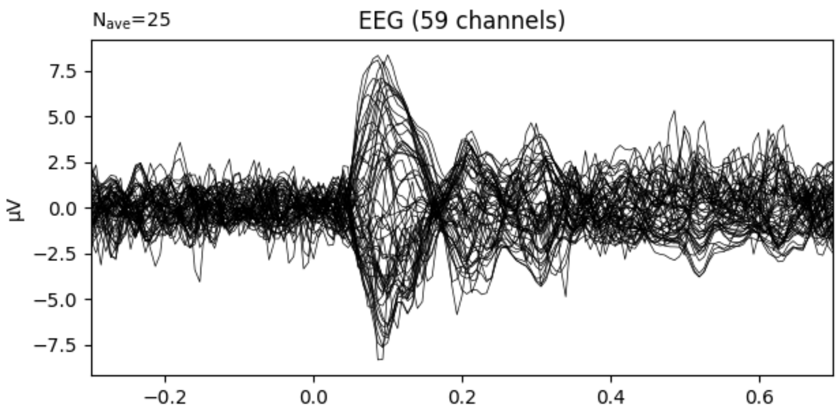

Figure 10 Audio Graph

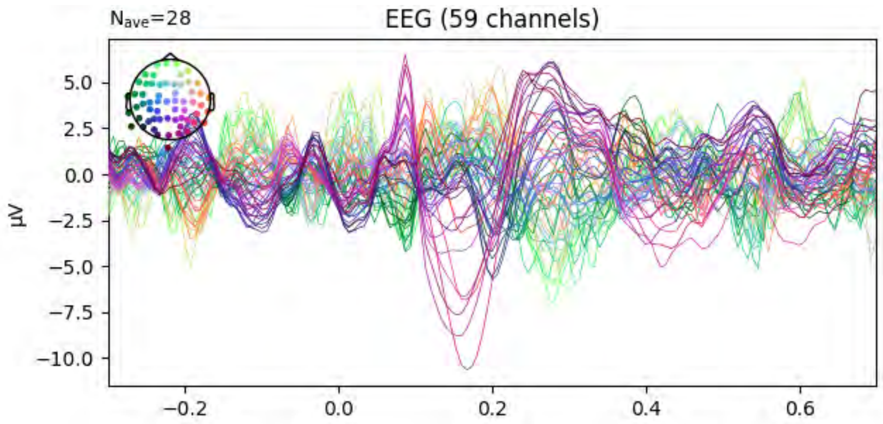

Figure 11 Video Graph

We will now use mne.viz.plot_compare_evokeds. This will combine all channels in each evoked object using global field power

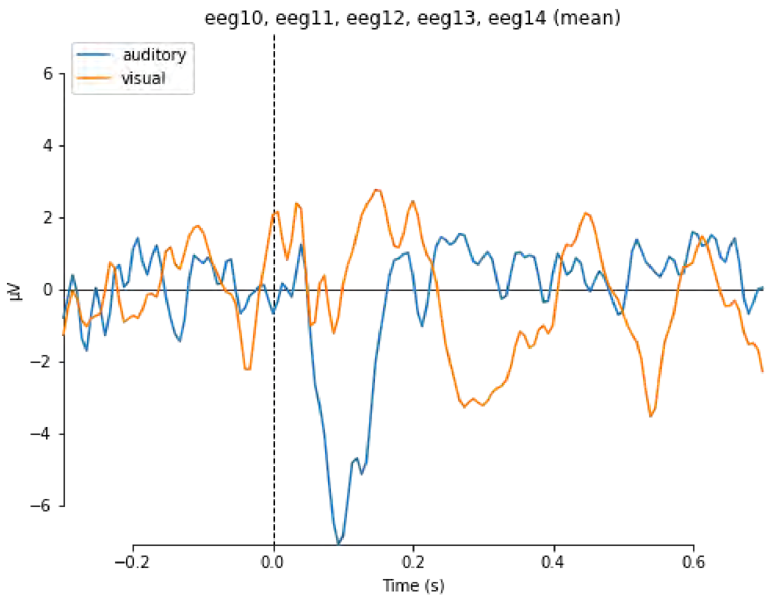

Figure 12 Comparison among audio and visual plots

Now we will find confidence levels, each epoch is treated alone

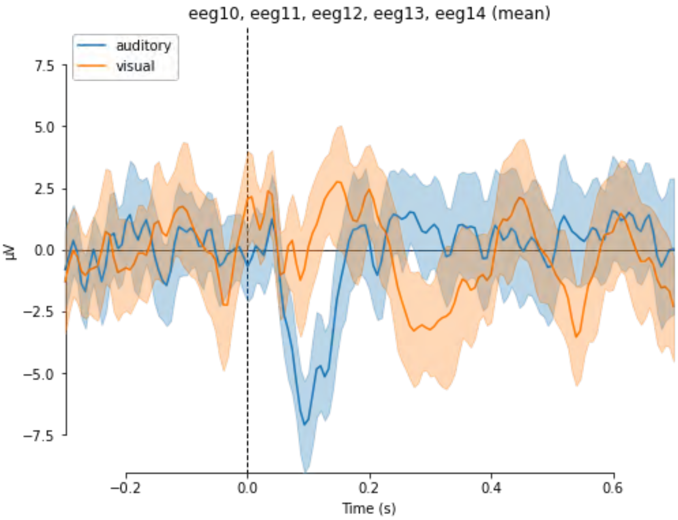

Figure 13 Confidence levels of auditory and visual plots

4.1.2 EMG data output Explanation:

As per the figure5 electrode placement, channel 1 and 2 is the EMG data of an eye movement during the speech generation. In, this research, mainly focused signals are on channel 6 which shows the muscle signals during the speech generation. This signals are used and connected with the tradition EL to check weather it is useful signals or not to develop an artificial EL.

In the figure, the pulse shown in the channel 6 are respiration pulses as subject is in the normal silent state. The repetation on this pulse are synchronous in silent mode which helps to trigger to turn off the device. However, channel 6 also helps to turn on the device but as a loss of 2 to 3 starting words of speech generation.

From this EMG signals, the electrode placement can be reduced to 4 to 5 electrodes as channel 6 is key focus for the triggering signals in this reseach. However, it is still unsufficient to work around due to difficulties in real time processing on EMG data.

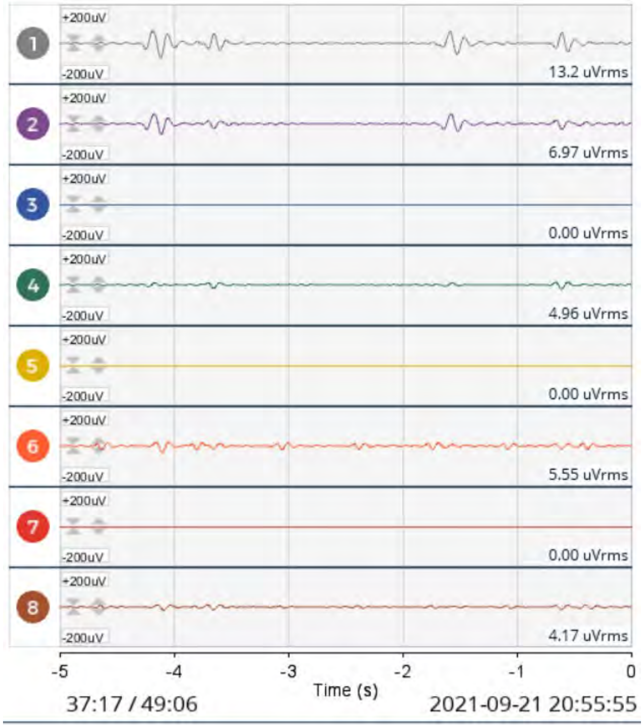

Figure 14 EMG output while speaking

4.2 Discussion

According to this literature, three of the main TLspeech rehabilitation methods differ from healthy speech both acoustically and perceptually. It is not the purpose of PROs to make comparisons between substitute speech rehabilitation groups and healthy speakers. With respect to all three acoustic parameters, fundamental frequency, maximum phonation time, and intensity, TES showed significantly better results than ES. According to TES, voice quality and intelligibility are significantly better than in ES and ELS. There were no obvious improvements in patient-reported outcomes among the speech rehabilitation groups. This is important to keep in mind. There is a low risk of bias in only three of the 26 included studies (level A). The most significant findings result from studies with a level B rating. Small numbers of patients are included in the included studies, and inferential statistics are not always performed. Acoustic measurements are often not specified in most studies, which may result in incorrect results. Our analysis of F0 and shimmer revealed several extreme outliers we had to exclude as a result [31, 38, 44]. Intensity measurements are acknowledged as being difficult, but only outcomes from individual studies are reported. It was found that standard measurement tools should be developed and used for evaluating substitute voice speakers in this systematic review. Evaluations of voice and speech are frequently considered gold standards for auditory-perceptual evaluations. Even so, it is important to acknowledge that there is great variability in the ratings. The development of rating schemes has been proposed by researchers [7, 16, 18]. However, they have not been widely adopted yet. New approaches to providing objective outcomes are being developed, with some promising results being reported recently [56, 57]. In our view, obtaining objective voice outcomes through automatic assessment tools may be the most promising way to analyze substitute voices, even though not all present tools seem suitable for doing so.

This study could only include a limited number of PRO studies. EORTC QLQ-H&N35 and EORTC QLQ-C30 are considered relevant outcome measures, but they were not included in the included studies. In our search for studies reporting the results of these questionnaires on the speech domain for various speaker groups, we did not find any. VHIs and V-RQOLs are usually used to assess vocal function after TL. Initially, the Communication and Participation Item Bank (CPIB) was not defined as a potential outcome of interest because it is a recently developed questionnaire [41]. A level A rated study of Eadie et al. [41] found a strong correlation between CPIB short form and VHI-10 scores. During speech rehabilitation, participants were asked to rate their own voice quality and intelligibility. These results were strongly related to CPIB short form scores as well. Accordingly, the CPIB short form has been found to be useful to elicit patients' opinions regarding vocal performance within the framework of the internationally recognized International Classification of Functioning (ICF) [38]. In comparison to ES, TES produces better results on the acoustic variables F0, MPT and intensity. Both methods of speech generation use the segment as a sound source. Acoustic voice outcomes might be more favorable with TES since it's a pulmonary-driven procedure. The tidal volume (roughly 5–600 ml) of TES may create a more steady and controlled airflow with the pulmonary airflow. High pressure may induce controlled hypertonicity or cranial positioning of the PE segment. F0 values in TES may be higher as a result of this. ES only provides a minimal amount of air, about 60–80 ml, which is around 2% of your lung capacity, and you are not able to control the pressure [1]. With such limited airflow and volume, ES sounds shorter and has a lower F0 and intensity. There was no date restriction on publication. The inclusion of older evidence is highly unlikely. When ES first came on the scene in the 1980s, it was considered to be the gold standard in speech rehab. In these years, ES probably had a decent education. The early publication period may have been better for esophageal speakers than the present.

CHAPTER 5: Conclusion & Future Work

5.1 Conclusion

Here, we describe a system for recognizing voice sounds based on myoelectric signals. We tested a variety of signal normalization and model adaptation methods to address the challenge of session dependence inherent to surface electromyography speech recognition. As a result of our study, we suggest that session adaptation in speech recognizers based on EEG signals can be achieved using methods used in conventional speech recognition systems [17]. When training data are shared between sessions, methods such as Variance Normalization and Maximum Likelihood adaptation are applied to improve across-session performance. Using seven EEG channels for within-session testing, we achieved a word accuracy of 97.3%. Across sessions, our recognition rates averaged 76.2%. By normalizing and adapting, we improved recognition rates to 87.1%. In comparison to other experiments, using more than two electrodes leads to significant performance improvement. The muscle movements that correspond to non-audible and audible speech show significant differences in our experiments. The recognition performance of our recognizer is slightly better when using audible speech, although combining training data can enhance the robustness of the resulting recognizer [23]. Using audible speech muscle signals, device can be turn on in **15.78 MicroSeconds** which cause loss of initial words to generate using EL. However, it only takes **7MicroSeconds** to turn off the device.

5.2 Future Work

Firstly, Improving the techniques of analyzing and processing of EMG signals to provide accurate result to vibrating circuit of the aftrificial larynx to turn on the device accurately on the start of the speech recognization or 2 to 3 MicroSeconds before speaking, So vowels cannot be missed.

Secondly, To work on minimizing the number of electrode placement on face and/or in order to avoid electrodes sticking to the user's face, robust non-contact sensors are needed.

Third, Improving the design and make it convenient to user friendly. Also, need to experiment on combining 2to3 different frequencies to generate the similar voice as human using an Artificial Electro-Larynx. It is also essential to move beyond discrete speech recognition to large vocabulary tasks which are continuously spoken [23-27], to improve the quality of spoken words.

References

1. Bien S, Rinaldo A, Silver CE, et al. History of voice rehabilitation following laryngectomy. Laryngoscope. 2008;118(3):453–458.

2. Czermak J. Uber die Sprache bei luftdichter Verschliessung des Kehlkopfs. [Concerning speech production of air-tight closure of the larynx]. Sitzungsberichte der kaiserlichen Academie der Wissenschafter mathematisch-naturwissenschaftliche Classe Wien; 1859. German.

3. Weir N. Otolaryngology: An Illustrated History. London, UK: Butterworths; 1990.

4. Gussenbauer C, Billroth T. Uber die erste durch Theodor Billroth am Menschen ausgeführte Kehlkopf-Exstirpation und die Anwendung des künstlichen Kehlkopfes. [On the first laryngeal extirpation carried out by Theodor Billroth in man, and the application of the artificial larynx]. Arch Klin Chirugie. 1874;(17):343–356. German.

5. Struebbing P. Pseudostimme nach Ausschaltung des Kehlkopfs, speziell nach Extirpation desselben. [Pseudostimulus after elimination of the larynx, especially after extirpation]. Dtsch Med Wochenshrift. 1888;14:1061. German.

6. Strubing P, D L. Erzeugung einer (natuerlichen) Pseudo-Stimme bei einem Manne mit totaler Extirpation des Kehlkopfes. [Generation of a (natural) pseudo-voice in a man with total extirpation of the larynx]. Arch Klin Chir. 1889;38:142. German.

7. Lowry LD. Artificial larynges: a review and development of a prototype self-contained intra-oral artificial larynx. Laryngoscope. 1981;91(8): 1332–1355.

8. Merriam-Webster Dictionary. Merriam-Webster Dictionary; 2016. Available from:http://www.merriam-webster. com/dictionary/sulcus. Accessed December 1, 2016.

9. Elmiyeh B, Dwivedi RC, Jallali N, et al. Surgical voice restoration after total laryngectomy: an overview. Indian J Cancer. 2010;47(3):239–247.

10. Perry AR, Shaw MA, Cotton S. An evaluation of functional outcomes (speech, swallowing) in patients attending speech pathology after head and neck cancer treatment(s): results and analysis at 12 months postintervention. J Laryngol Otol. 2003;117(5):368–381.

11. Bohnenkamp TA, Stowell T, Hesse J, Wright S. Speech breathing in speakers who use an electrolarynx. J Commun Disord. 2010;43(3): 199–211.

12. Liu H, Wan M, Wang S, Niu H. Aerodynamic characteristics of laryngectomees breathing quietly and speaking with the electrolarynx. J Voice. 2004;18(4):567–577.

13. Babin E, Beynier D, Le Gall D, Hitier M. Psychosocial quality of life in patients after total laryngectomy. Rev Laryngol Otol Rhinol (Bord). 2009;130(1):29–34.

14. Singer S, Danker H, Dietz A, et al. Sexual problems after total or partial laryngectomy. Laryngoscope. 2008;118(12):2218–2224.

15. Varghese BT, Mathew A, Sebastian P, Iype EM, Vijay A. Comparison of quality of life between voice rehabilitated and nonrehabilitated laryngectomies in a developing world community. Acta Otolaryngol. 2011;131(3):310–315.

16. Finizia C, Bergman B. Health-related quality of life in patients with laryngeal cancer: a post-treatment comparison of different modes of communication. Laryngoscope. 2001;111(5):918–923.

17. Eadie TL, Doyle PC. Quality of life in male tracheoesophageal (TE) speakers. J Rehabil Res Dev. 2005;42(1):115–124.

18. Eadie TL, Day AM, Sawin DE, Lamvik K, Doyle PC. Auditoryperceptual speech outcomes and quality of life after total laryngectomy. Otolaryngol Head Neck Surg. 2013;148(1):82–88.

19. Liu H, Ng ML. Electrolarynx in voice rehabilitation. Auris Nasus Larynx. 2007;34(3):327–332.

20. Bloom E. The artificial larynx: past and present. In: Salmon SJ, Goldstein LP, editors. The Artificial Larynx Handbook. New York, NY: Grune & Stratton; 1978:57–86.

21. Williams WG, Ostroy L. A modified electrolarynx for laryngectomees. J Clin Eng. 1977;2(2):149–152.

22. Goldstein EA, Heaton JT, Stepp CE, Hillman RE. Training effects on speech production using a hands-free electromyographically controlled electrolarynx. J Speech Lang Hear Res. 2007;50(2):335–351.

23. Wang L, Feng Y, Yang Z, Niu H. Development and evaluation of wheelcontrolled pitch-adjustable electrolarynx. Med Biol Eng Comput. Epub 2016 Dec 24.

24. Saikachi Y, Stevens KN, Hillman RE. Development and perceptual evaluation of amplitude-based F0 control in electrolarynx speech. J Speech Lang Hear Res. 2009;52(5):1360–1369.

25. Espy-Wilson CY, Chari VR, MacAuslan JM, Huang CB, Walsh MJ. Enhancement of electrolaryngeal speech by adaptive filtering. J Speech Lang Hear Res. 1998;41(6):1253–1264.

26. Niu HJ, Wan MX, Wang SP, Liu HJ. Enhancement of electrolarynx speech using adaptive noise cancelling based on independent component analysis. Med Biol Eng Comput. 2003;41(6):670–678.

27. Watson PJ, Schlauch RS. The effect of fundamental frequency on the intelligibility of speech with flattened intonation contours. Am J Speech Lang Pathol. 2008;17(4):348–355.

28. Choi HS, Park YJ, Lee SM, Kim KM. Functional characteristics of a new electrolarynx "Evada" having a force sensing resistor sensor. J Voice. 2001;15(4):592–599.

29. Uemi N, Ifukube T, Takahashi M, Matsushima J. Design of a new electrolarynx having a pitch control function. Paper presented at: 3rd IEEE International Workshop on Robot and Human Communication (RO-MAN '94); 1994; Nagoya, Japan.

30. Liu H, Wan M, Ng ML, Wang S, Lu C. Tonal perceptions in normal laryngeal, esophageal, and electrolaryngeal speech of Mandarin. Folia Phoniatr Logop. 2006;58(5):340–352.

31. Takahashi H, Nakao M, Kikuchi Y, Kaga K. Alaryngeal speech aid using an intra-oral electrolarynx and a miniature fingertip switch. Auris Nasus Larynx. 2005;32(2):157–162.

32. Takahashi H NM, Okuas T, Hatamura Y, Kikuchi Y, Kaga K. A voicegeneration system using an intra-mouth vibrator. J Artif Organs. 2001;4:288–294.

33. Staffieri A, Mostafea BE, Varghese BT, et al. Cost of tracheoesophageal prostheses in developing countries. Facing the problem from an internal perspective. Acta Otolaryngol. 2006;126(1):4–9.

34. Xi S. Effectiveness of voice rehabilitation on vocalisation in postlaryngectomy patients: a systematic review. Int J Evid Based Healthc. 2010;8(4):256–258.

35. Singer MI, Blom ED. An endoscopic technique for restoration of voice after laryngectomy. Ann Otol Rhinol Laryngol. 1980;89(6 Pt 1):529–533.

36. Sinclair CF, Rosenthal EL, McColloch NL, et al. Primary versus delayed tracheoesophageal puncture for laryngopharyngectomy with free flap reconstruction. Laryngoscope. 2011;121(7):1436–1440.

37. Makitie AA, Niemensivu R, Juvas A, Aaltonen LM, Back L, Lehtonen H. Postlaryngectomy voice restoration using a voice prosthesis: a single institution's ten-year experience. Ann Otol Rhinol Laryngol. 2003;112(12):1007–1010.

38. Clements KS, Rassekh CH, Seikaly H, Hokanson JA, Calhoun KH. Communication after laryngectomy. An assessment of patient satisfaction. Arch Otolaryngol Head Neck Surg. 1997;123(5):493–496.

39. Ward EC, Koh SK, Frisby J, Hodge R. Differential modes of alaryngeal communication and long-term voice outcomes following pharyngolaryngectomy and laryngectomy. Folia Phoniatr Logop. 2003;55(1): 39–49.

40. Weiss MS, Basili AG. Electrolaryngeal speech produced by laryngectomized subjects: perceptual characteristics. J Speech Hear Res. 1985;28(2): 294–300.

41. Siric L, Sos D, Rosso M, Stevanovic S. Objective assessment of tracheoesophageal and esophageal speech using acoustic analysis of voice. Coll Antropol. 2012;36(Suppl 2):111–114.

42. Fagan MJ, Ell SR, Gilbert JM, Sarrazin E, Chapman PM. Development of a (silent) speech recognition system for patients following laryngectomy. Med Eng Phys. 2008;30(4):419–425.

43. Gilbert JM, Rybchenko SI, Hofe R, et al. Isolated word recognition of silent speech using magnetic implants and sensors. Med Eng Phys. 2010;32(10):1189–1197.

44. Kohlberg GD, Gal YK, Lalwani AK. Development of a Low-Cost, Noninvasive, Portable Visual Speech Recognition Program. Ann Otol Rhinol Laryngol. 2016;125(9):752–757.

Appendix

Research History

A laryngeal prosthesis has been used to restore voice for over 150 years. Among the first described laryngeal prostheses was that of Czermak 1859, who described voice production in a laryngeal stenotic patient by diverting the airflow from a tracheal cannula to the mouth with a reed-filled tube. A reed-like pneumatic device was mounted on the end of a double lumen tracheostomy tube at the pharyngectomy of Billroth's first laryngectomy in 1873 and provided with the speech apparatus by Gussenbauer. The successful outcome was announced at the German Society of Surgeons Third Congress in 1874. Voice restoration progressed rapidly in the late 19th century with reports of esophageal speech, mechanical vibrations in the early 1900s, fistulas that allowed air to pass into the pharynx and upper esophagus during the middle of the 20th century, and tracheoesophageal puncture (TEP) speech in the mid-20th century that involved a valve which inserted into the trachea and upper esophagus.

When the larynx is surgically removed due to conditions such as laryngeal cancer, a person is unable to speak naturally [4]. Three methods are used to restore the voicing without the use of vocal chords and the space between them - Esophageal speech, Tracheoesophageal speech, and electrolarynx. In esophageal speech, air is introduced by swallowing the air and then released abruptly from the esophagus to the oral cavity [5]. Pharryngeal muscles vibrate as a result. An articulator converts this vibration to sound. Practicing this skill is imperative. A one-way valve placed on a tracheal puncture allows a patient to take in air for tracheoesophageal speech [7]. In order to produce speech, the articulators occlude the valve to direct the air to the oral cavity. There are several hygienic issues presented by this method, including fungal infections leading to fluid leaks through tracheo-esophageal punctures [8]. To reestablish speaking, the electrolarynx replaces the larynx with an electromechanical vibrator. In order to speak, the device could be held against the neck with a waveform generator and vibrating head. As the device vibrates, acoustic energy is released and the movement of the articulators produces speech. A patient who is ill and under artificial ventilation may also benefit from the electrolarynx [10-12]. Electrolarynxes are known for their conspicuous appearance. Since the electrolarynx is so large, it must be hand held throughout verbal communication, which would cause a great deal of inconvenience and awkwardness to the user. Researchers are considering several ways to make patient's lives easier through miniaturization. An attachment brace with a thin vibrator [13] makes it possible to attach the device to the neck surface. Transducer control is via a small wireless controller that fits in the pocket. Even with a 9-volt supply required, the whole system weighs a lot. YourTONE II, an electrolarynx, is wearable but the size and weight of the device aren't revealed in the literature [14]. Using a tiny pager motor, an innovative hands free operation design [15] was developed. When a voltage is applied to the motor, a thin polyethylene membrane pulsates. In general, pager motors have low torque handling capacity, resulting in insufficient vibration to produce audible speech.

The PE segment is driven by mucosal vibrations caused by pulmonary air in TES. An electrolarynx, a sound-producing device, mostly handheld, can be placed against one's neck or cheek, thus replacing the external sound source in ELS (Electrolarynx Sound System). The best speech rehabilitation method for restoring oral communication is not agreed upon worldwide based on science. According to some theories, TL patients who have better voice quality will also enjoy better quality of life [2, 3]. Multidimensional assessment is recommended to evaluate speech rehabilitation outcomes [4, 5]. Among the three substitute speech options, this systematic review compares acoustics, perception, and patient-reported outcomes (PROs). The pitch and amplitude of a voice are regularly measured in acoustic voice analysis [6]. Standard acoustic voice analysis does not always work when it comes to measuring substitute voices, which are characterized by having more noise components and less regularity than laryngeal voices [7].

In addition to the deviances in regularity compared to laryngeal voices, sensory evaluations of speech rehabilitation methods require a well-considered approach. The most suitable methods for evaluating substitute voices are to evaluate the quality of the voice and the intelligibility of the spoken word [8, 9]. The impact of speech rehabilitative treatments is typically evaluated with Quality of Life (QOL) questionnaires such as the EORTC QLQ-H&N35 and/or the EORTC QLQ-C30, which has questions about speech functioning [10, 11]. Speech rehabilitation results are better understood with PROs, such as the Voice Handicap Index (VHI) or Voice-Related Quality of Life (V-RQOL) [10–14].

A person's ability to speak naturally is terminated when the larynx is removed surgically as a result of laryngeal cancer [4]. Esophageal speech, tracheoesophageal speech, and electrolarynx are the three methods of re-establishing voicing without vocal chords and their space. By swallowing air, a person introduces air into the esophagus region and releases it abruptly into the oral cavity during esophageal speech [5]. Pharyngeal muscles vibrate as a result. The articulators convert these vibrations into speech. It requires a great deal of practice and training

[6]. The one-way valve attached to a tracheal puncture is used in tracheo-esophageal speech [7]. By occluding the valve, the articulators direct air to the oral cavity, where it forms speech. In this method, there are hygienic problems such as fungal infections that can cause fluid leaks through tracheoesophageal puncture [8]. The electrolarynx is an electromechanical vibrator that replaces the larynx so that speech can be generated. During speaking, the device is held against the neck and a waveform generator generates a vibration. During speech production, the articulators move in response to vibrations of the device, converting vibrational energy into acoustic energy [9]. Electrolarynxes are also helpful for patients undergoing artificial ventilation when they are ill [10-12].

In addition to its conspicuous appearance, electrolarynxes have some other disadvantage The electrolarynx is a large device, which requires that the user hold it throughout verbal communication, causing inconveniences and awkwardness for the user. To make patients' lives easier, researchers are considering several miniaturization concepts. With the goal of reducing the size of the device, we designed a thin vibrator [13] that can be attached to the surface of the neck through the use of a brace. Using the wireless controller, the transducer can be controlled. The entire system is still heavy due to the requirement of 9 V supply. The wearable electrolarynx YOUR TONE II does not reveal the size or weight of its electrolarynx [14]. The motor of a tiny pager was used to implement a hands-free design [15]. In this application, the motor is attached to a thin membrane that pulses when voltage is applied. Pager motors have an insufficient handling capacity, so the vibration generated during speech is not audible. Speech intelligibility is affected by loudness reduction [16, 17]. Using a video camera and a tiny transducer, the hands-free design approach controls lip movement to enable electrolarynx control; however, it is not yet known whether the voice is audible [18]. Using mechanically driven gears, the artificial larynx has a fundamental frequency range of under 100 Hz [19]. This would result in a voice that is distracted from what is being spoken.

The current speech rehabilitation options have not been subject to a comprehensive review of the pros and cons. The collection of the best available evidence regarding the three speech rehabilitation methods would likely lead to a consensus as to which speech rehabilitation to use after TL and could assist clinicians, patients, and reimbursement agencies in making decisions. We investigated the acoustic, perceptual, and PRO effects of the three speech rehabilitation techniques following TL in this systematic review. In this research, we will investigate how the outcomes of various speech rehabilitation methods compare to those of normal laryngeal speech (healthy speakers), as well as what types of results are most favorable for each rehabilitation method. An examination of the literature on the outcome of speech following total laryngectomy (TL) was conducted using a systematic search strategy. In this search strategy, we focused specifically on the primary and secondary results that we were looking for. Depending on the literature, we selected the best primary and secondary results. The objective of the acoustic outcomes was to elucidate options for speech rehabilitation from objective data. Perceptual ratings and PROs served as vehicles for obtaining subjective information about the voices. In order to identify primary acoustic outcomes, we have selected fundamental frequency (Fo), harmonic to noise ratio (HNR), and voicedness percentage (%VO). Numerous authors have indicated that these outcomes are crucial to determining pitch, stability, and noise characteristics [7, 15, 16–17]. Other acoustic outcomes, such as jitter, shimmer, intensity, spectral tilt, and maximum phonation time (MPT), were interesting. The literature uses many of these outcome variables, although some are not as reliable in substitute voicing [16, 17].

The IINFVo scale was used to assess impression, intelligibility, noise, fluency, and voice quality, which are basic perceptual outcomes of interest. In addition to GRAAS, secondary perceptual outcomes of relevance were chosen from well-established perceptual assessment tools, such as unintended additive noise, fluency, and voicing functions [8, 18], and other recommended perceptual parameters of TL-speech in the literature. Among the most popular PROs are VHI13 and V-RQOL14. In addition, we included communication specific PROs on the EORTC QLQ-H&N35 [11] and the EORTC QLQ-C30 [10], which evaluate general quality of life including subsets related to communication.

Restoration of Voice History

Over 150 years ago, Czermak reported voice production in a patient with complete laryngeal stenosis when airflow was diverted through a reed tube from a tracheal cannula through the mouth. The patient's assistant created a custom speech apparatus consisting of a tracheostomy tube with a double lumen and an inlet extending into the pharyngostome for him, which was mounted onto a pneumatic device. In 1874, the German Company of Surgeons reported the successful result at its Third Congress in Berlin. Voice recovery advanced rapidly with esophageal speech reported in the mid-19th century, mechanical vibrations at the turn of the 19th century, and air conduits that enabled upper esophagus and pharynx to be reached in the mid-20th century, and tracheoesophageal puncture (TEP) speech that used bidirectional prosthetic valves in the mid-20th century [22][26].

Code for EEG

```
import os

import numpy as np

import matplotlib.pyplot as plt

import mne


sample_data_folder = mne.datasets.sample.data_path()

sample_data_raw_file = os.path.join(sample_data_folder, 'MEG', 'sample',

                                    'sample_audvis_filt-0-40_raw.fif')

raw = mne.io.read_raw_fif(sample_data_raw_file, preload=False)


sample_data_events_file = os.path.join(sample_data_folder, 'MEG', 'sample',

                                       'sample_audvis_filt-0-40_raw-eve.fif')

events = mne.read_events(sample_data_events_file)


raw.crop(tmax=90)  # in seconds; happens in-place

# discard events >90 seconds (not strictly necessary: avoids some warnings)

events = events[events[:, 0] <= raw.last_samp]\

channel_renaming_dict = {name: name.replace(' 0', '').lower()

                         for name in raw.ch_names}

_ = raw.rename_channels(channel_renaming_dict)
```

```
raw.plot_sensors(show_names=True)

fig = raw.plot_sensors('3d')

for proj in (False, True):

    fig = raw.plot(n_channels=5, proj=proj, scalings=dict(eeg=50e-6))

    fig.subplots_adjust(top=0.9)  # make room for title

    ref = 'Average' if proj else 'No'

    fig.suptitle(f'{ref} reference', size='xx-large', weight='bold')

raw.filter(l_freq=0.1, h_freq=None)

np.unique(events[:, -1])

event_dict = {'auditory/left': 1, 'auditory/right': 2, 'visual/left': 3,

              'visual/right': 4, 'face': 5, 'buttonpress': 32}

epochs = mne.Epochs(raw, events, event_id=event_dict, tmin=-0.3, tmax=0.7,

                    preload=True)

fig = epochs.plot()

reject_criteria = dict(eeg=100e-6,  # 100 µV

                       eog=200e-6)  # 200 µV

_ = epochs.drop_bad(reject=reject_criteria)

epochs.plot_drop_log()

l_aud = epochs['auditory/left'].average()

l_vis = epochs['visual/left'].average()

fig1 = l_aud.plot()

fig2 = l_vis.plot(spatial_colors=True)

l_aud.plot_topomap(times=[-0.2, 0.1, 0.4], average=0.05)

l_aud.plot_joint()
```

```
#GFP Calculation

for evk in (l_aud, l_vis):

    evk.plot(gfp=True, spatial_colors=True, ylim=dict(eeg=[-12, 12]))

l_aud.plot(gfp='only')

gfp = l_aud.data.std(axis=0, ddof=0)


# Reproducing the MNE-Python plot style seen above

fig, ax = plt.subplots()

ax.plot(l_aud.times, gfp * 1e6, color='lime')

ax.fill_between(l_aud.times, gfp * 1e6, color='lime', alpha=0.2)

ax.set(xlabel='Time (s)', ylabel='GFP (µV)', title='EEG')

left = ['eeg17', 'eeg18', 'eeg25', 'eeg26']

right = ['eeg23', 'eeg24', 'eeg34', 'eeg35']


left_ix = mne.pick_channels(l_aud.info['ch_names'], include=left)

right_ix = mne.pick_channels(l_aud.info['ch_names'], include=right)

roi_dict = dict(left_ROI=left_ix, right_ROI=right_ix)

roi_evoked = mne.channels.combine_channels(l_aud, roi_dict, method='mean')

print(roi_evoked.info['ch_names'])

roi_evoked.plot()

evokeds = dict(auditory=l_aud, visual=l_vis)

picks = [f'eeg{n}' for n in range(10, 15)]

mne.viz.plot_compare_evokeds(evokeds, picks=picks, combine='mean')

evokeds = dict(auditory=list(epochs['auditory/left'].iter_evoked()),
```

```
            visual=list(epochs['visual/left'].iter_evoked()))

mne.viz.plot_compare_evokeds(evokeds, combine='mean', picks=picks)

aud_minus_vis = mne.combine_evoked([l_aud, l_vis], weights=[1, -1])

aud_minus_vis.plot_joint()

grand_average = mne.grand_average([l_aud, l_vis])

print(grand_average)

list(event_dict)

epochs['auditory'].average()
```

NLTK Coding

```
import nltk

nltk.download()

def unigram_features (words):

   """

   This is the simplest possible feature representation of a document.



   Each word is a feature.

   """

   return dict((word, True) for word in words)




def extract_features (corpus, file_ids, cls, feature_extractor=unigram_features):

   """

   Turn a set of files all belonging to one class into a list

   of (feature dictionary, cls) pairs, to be used in testing or training
```

```
    a classifier.
    """
    return [(feature_extractor(corpus.words(i)), cls) for i in file_ids]



def get_words_from_corpus (corpus, file_ids):


    for file_id in file_ids:
        words = corpus.words(file_id)
        for word in words:
            yield word
# Using a corpus of movie review data
# 2000 positive and negative reviews, evenly balanced.
from nltk.corpus import movie_reviews as mr
data = dict(pos = mr.fileids('pos'),
            neg = mr.fileids('neg'))
print mr.raw(data['pos'][0])[:100]
from nltk.corpus import movie_reviews as mr

# Use a Naive Bayes Classifier
from nltk.classify import NaiveBayesClassifier

data = dict(pos = mr.fileids('pos'),
            neg = mr.fileids('neg'))
```

```
##########################################################################
#
#  Dividing up the data
##########################################################################

# Use 90% of the data for training
test_start_index = 900

neg_training = extract_features(mr, data['neg'][:test_start_index], 'neg',
                                feature_extractor=unigram_features)

# Use 10% for testing the classifier on unseen data.
neg_test = extract_features(mr, data['neg'][test_start_index:], 'neg',
                                feature_extractor=unigram_features)

pos_training = extract_features(mr, data['pos'][:test_start_index],'pos',
                                feature_extractor=unigram_features)

pos_test = extract_features(mr, data['pos'][test_start_index:],'pos',
                                feature_extractor=unigram_features)

train_set = pos_training + neg_training
```

```
test_set = pos_test + neg_test

classifier = NaiveBayesClassifier.train(train_set)

def get_review_text (clf,file_id,start=0,end=None):

    words = list(mr.words(data[clf][file_id]))

    return ' '.join(words[start:end])


print get_review_text('pos',0,end=95)


print '     . . . . . . '


print get_review_text('pos',0,start=-190)

predicted_label0 = classifier.classify(pos_test[0][0])


print 'Predicted: %s Actual: pos' % (predicted_label0,)


print get_review_text('neg',0,end=120)


print '     . . . . . . '


print get_review_text('neg',0,start=-180)


predicted_label1 = classifier.classify(neg_test[0][0])


print 'Predicted: %s Actual: neg' % (predicted_label1,)
```

```
# To see the the feature dictionary passed in to the classifier,
# uncomment the next line
#pos_test[0][0]

from sklearn.metrics import precision_score, recall_score,accuracy_score

def do_evaluation (pairs, pos_label='pos', verbose=True):
    predicted, actual = zip(*pairs)
    (precision, recall,accuracy) = (precision_score(actual,predicted,pos_label=pos_label),
                                    recall_score(actual,predicted,pos_label=pos_label),
                                    accuracy_score(actual,predicted))
    if verbose:
        print_results(precision, recall, accuracy, pos_label)
    return (precision, recall,accuracy)

def print_results (precision, recall, accuracy, pos_label):
    banner =  'Evaluation with pos label = %s' % pos_label
    print
    print banner
    print '=' * len(banner)
    print '{0:10s} {1:.1f}'.format('Precision',precision*100)
    print '{0:10s} {1:.1f}'.format('Recall',recall*100)
    print '{0:10s} {1:.1f}'.format('Accuracy',accuracy*100)
```

```
pairs = [(classifier.classify(example), actual)
            for (example, actual) in test_set]


do_evaluation (pairs)

pos_guesses = [p for (p,a) in pairs if p=='pos']

pos_actual = [a for (p,a) in pairs if a=='pos']

do_evaluation (pairs, pos_label='neg')

print 'Note that {:.1%} of our classifier guesses were positive'.format(float(len(pos_guesses))/len(pairs))

print 'While {:.1%} of the reviews were actually positive'.format(float(len(pos_actual))/len(pairs))

# to see the actual pairs that came out of the test uncomment the next line

#pairs

#SVM

from sklearn.feature_extraction.text import TfidfVectorizer

from sklearn.svm import LinearSVC


import os.path


def add_data_from_files (file_list,data_list):

    for f in file_list:
```

```
        with open(f,'r') as fh:

            data_list.append(fh.read())


home = os.getenv('HOME')

# This is where MY NLTK data is.  Yours should be in a similar place relative

# to what your machine thinks is HOME.

data_dir = os.path.join(home,'nltk_data/corpora/movie_reviews/')


clses = ['pos','neg']


#  The data is in the data_dir, sorted into subdirectories, one for each class.

data_dirs = [os.path.join(data_dir,cls) for cls in clses]

#  We use a somewhat more traditional feature weights, called TFIDF weights

vectorizer = TfidfVectorizer(sublinear_tf=True, max_df=0.5,

                             stop_words='english')


# We're going to compute 4 lists training data and labels, test data a nd labels

train_labels = []

test_labels = []


train_data = []

test_data = []

training_proportion = (9,10)

for i,cls  in enumerate(clses):
```

```
    d_dir = data_dirs[i]
    os.chdir(d_dir)
    cls_files = os.listdir(d_dir)
    num_cls_files = len(cls_files)
    training_index = (training_proportion[0] *(num_cls_files/training_proportion[1]))
    train_labels.extend(cls for f in cls_files[:training_index])
    test_labels.extend(cls for f in cls_files[training_index:])
    add_data_from_files (cls_files[:training_index],train_data)
    add_data_from_files (cls_files[training_index:],test_data)

# Now with data set represented as a list of strings (one from each file),
# extract the TFIDF features
train_features = vectorizer.fit_transform(train_data)

#  We extract features from the test data using the same vectorizer
#  trained on training data. The TFIDF feature model has been fit to
#  (depends only on) the training data.
test_features = vectorizer.transform(test_data)

# Create an SVM classifier instance
clf = LinearSVC(loss='squared_hinge', penalty="l2",
            dual=False, tol=1e-3)

# Train (or "fit") the model to the training data.
```

```
clf.fit(train_features, train_labels)


# Test the model on the test data.
predicted_labels = clf.predict(test_features)


# Evaluate the results
pos_guesses = [p for p in predicted_labels if p=='pos']
pos_actual = [p for p in test_labels if p=='pos']
print 'Note that {:.1%} of our classifier guesses were positive'.format(float(len(pos_guesses))/len(test_labels))
print 'While {:.1%} of the reviews were actually positive'.format(float(len(pos_actual))/len(test_labels))
do_evaluation (zip(predicted_labels,test_labels), pos_label='pos', verbose=True)
do_evaluation (zip(predicted_labels,test_labels), pos_label= 'neg', verbose=True)
```

Figure removed due to copyright restriction

Figure 15 Traditional Electrolarynx

Figure removed due to copyright restriction

Figure 16 OpenBCI Cyton board

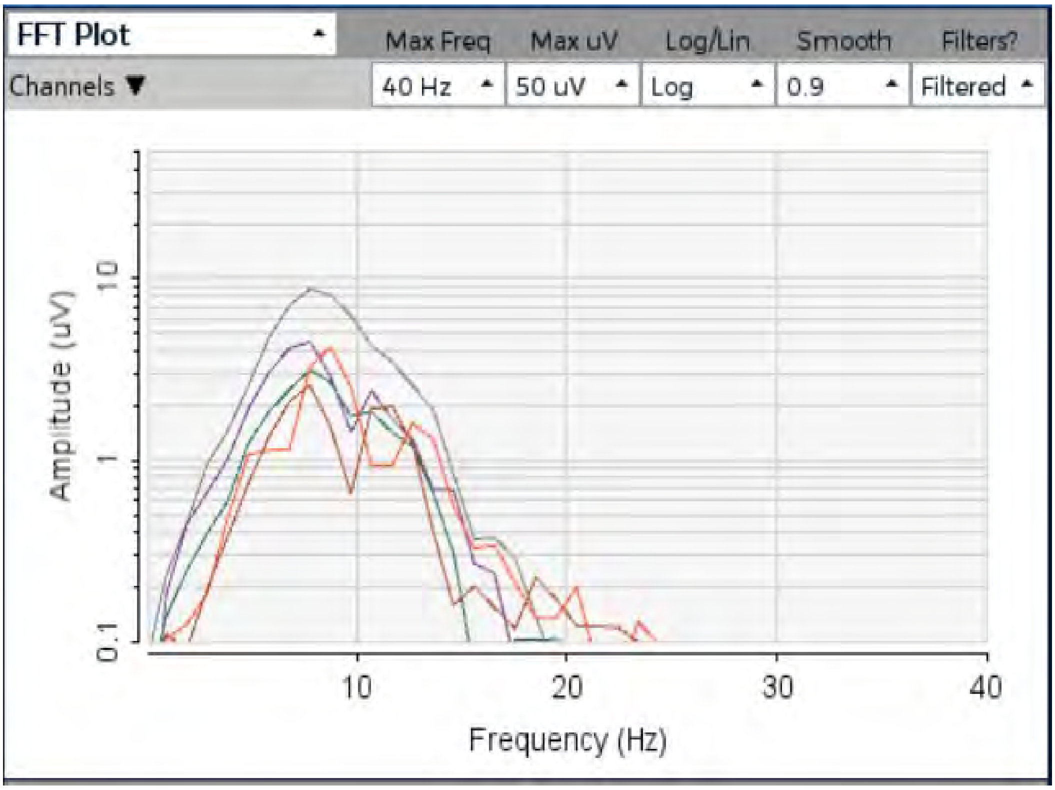

Figure 17 FFT plot of EMG signals in Fig(14)

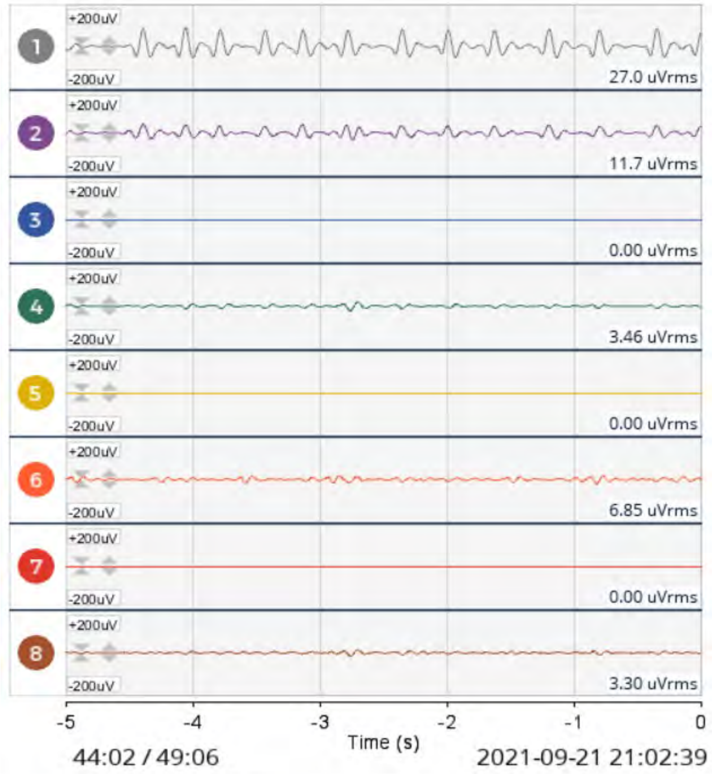

Figure 18 Constant Eye blinking EMG data

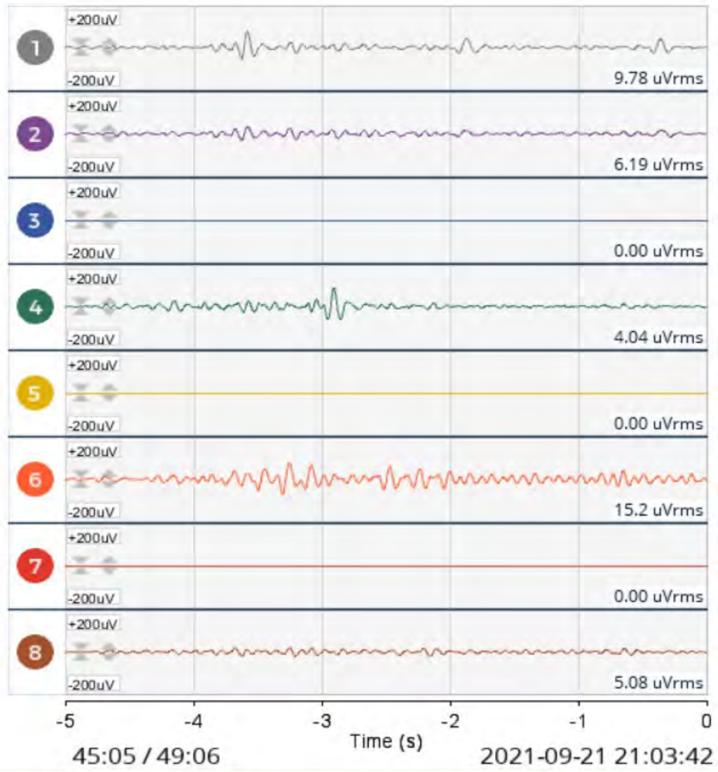

Figure 19 Constant speaking EMG data

Figure removed due to copyright restriction

Figure 20 Artificial Larynx design (Imaginary, not working)

Figure removed due to copyright restriction

Figure 21 Future dream if this concept work