Augmentative Communication Device

BME 301
Department of Biomedical Engineering
University of Wisconsin-Madison
May 5, 2010

Prachi Agarwal, BWIG
Erin Devine, Communicator
Brian Mogen, Team Leader
Steve Wyche, BSAC

Client
Dr. Lawrence Kaplan, M.D
Department of Pediatrics, UW School of Medicine and Public Health

Advisor
Dr. Amit Nimunkar
Department of Biomedical Engineering

Table of Contents

Abstract

The goal of this project is to develop a novel method of communication for individuals suffering from a wide range of communicative disorders. Currently, there are no devices on the market that are fast, inexpensive, aesthetic, and allow the user to add emotion to their communication. Our design provides instantaneous speech, interfaces with a number of sound sources, and attaches to the user's belt in an aesthetically-pleasing manner.

Background Information

Neuromotor dysfunction presents itself in a number of forms, one of the most common being cerebral palsy. This occurs in approximately 1 out of every 500 people and is a result of abnormalities in the growth and functioning of the brain[1]. This leads to uncontrollable reflex movements and moderate to severe muscle tightness. Cerebral palsy can be caused by head trauma after birth, but this is relatively rare. It is more common for the brain to be affected before or during birth.

Four main types of brain damage contribute to the majority of cerebral palsy cases[2]. The first is periventricular leukomalacia, which is damage to the white matter of the brain. This is usually responsible for transmitting signals throughout the brain and body, but small holes in this white matter that form before birth do not allow this to develop properly. Another cause of cerebral palsy is cerebral dysgenesis, or abnormal development of the brain. During the first 20 weeks of development, the fetal brain is very vulnerable. Any interruption in the growth of the brain causes abnormalities that interfere with the transmission of signals. Mutations in genes, infections, fevers, or trauma could contribute to this interruption. Intracranial hemorrhage, or bleeding in the brain, is also a possibility. If blood flow is blocked by blood clots in the placenta, the baby may suffer a stroke, leading to blocked or broken vessels in the brain. The final key development malfunction is hypoxic-ischemic encephalopathy, or intrapartum asphyxia. More commonly referred to simply as asphyxia, this is a lack of oxygen in the brain. Tissue in the brain, most notably in the cerebral motor cortex, can be destroyed, and this causes cerebral palsy.

Motor functions are affected differently in everyone; some have a slight limp, while others are completely wheelchair-bound. Those with spastic hemiplegia are mostly affected in the arms and hands; those with spastic diplegia are more affected in the legs and feet[3]. The most severe form is spastic quadriplegia, where one has severe stiffness in the limbs, is usually completely wheelchair-bound, and has extreme difficulties speaking. Cerebral palsy is a non-progressive disorder, meaning the disease will not worsen, but later physiological disabilities are very common. Many of these neurological conditions cause dysarthia which impairs a person's ability to speak[6].

Laryngeal cancer is not as common as cerebral palsy, only affecting about 1 in 22,666 people. This is a disease in which malignant cancer cells develop in the tissues of the larynx[4]. Sometimes called the "voice box", the larynx houses the vocal cords and is found just below the pharynx in the neck. When a person attempts to speak, the vocal cords vibrate as air moves against them, producing sound. This sound echoes through the person's mouth and nose to create a voice. The cancer typically develops in the squamous cells, which line the inside of the larynx. To treat the laryngeal cancer, many people choose to have a laryngectomy. In this procedure, the larynx is surgically removed. Since the vocal cords are located within the larynx, these are also removed. This causes the affected individual to lose all speech capabilities.

Finally, speech can be inhibited with a paralyzed diaphragm. The diaphragm is vital to normal respiration. In normal speech, air is required to be pushed up and out of the body. The diaphragm serves the purpose of pushing this air. When the diaphragm is paralyzed, the individual is not able to produce sufficient air flow to generate normal speech[7].

Existing Devices

The devices currently available on the market today consist mostly of touch screen tablet PCs or handheld devices. These devices have pre-programmed common phrases and keyboards to enter in custom sentences. In order to give the user the ability to speak more quickly, the devices usually prompt possibilities for the next letter or word, but typing what they want to say is still a laborious task and is frustrating because it is time consuming.

One such device is the Tango. It is geared towards children and uses images and icons to direct the child to what they want to say. This device, however, is limited only to pre-programmed phrases. While useful for kids who can’t speak, this device stifles a child’s creativity in that they cannot produce their own sentences. It does not allow children to expand their vocabulary and limits them to what is programmed into the device.

Another device on the market is the Dasher. This device uses some sort of pointer, whether it is a joystick, a mouse type apparatus or a slider, to point out letters to form words.

The program prompts the user with possible and common letters to follow the first in order to speed up the input process. While this is a great interface for someone that cannot use a standard keyboard to type, this is still very slow and makes the user less likely to say something that would be time consuming. In our client's experience, people then tend to limit their speech and vocabulary to the minimum that is required for what they want to communicate. In this way, devices like this limit self-expression in addition to being unable to add inflection or emphasis to the words.

Another manufacturer, DynaVox, produces touch screen devices that have some pre-programmed common phrases and a keypad input system. This interface is also slow, cumbersome, and lacks the ability to add emotion. While these devices come in small handheld versions or full size tablet PC based on the preference of the user, they are still limiting because they are slow and users encounter the same frustration as they do with other devices. Users pay a lot of money for these expensive devices, and then tend to not use them because they are not an efficient means of communication.

Problem Statement

As a continuing project, our goal is to move the project forward in a direction that will help us on the way to get a patent. In a preliminary meeting with WARF, they requested that we identify what about our device is patentable. The main goal for this semester is to characterize something about the device that is novel, unique and cannot be designed around by any possible competitors. In addition, we would like to finish the semester with a working prototype that could feasibly be used by an individual in need of speech assistance.

Physiology

The physiology behind speech in a healthy involves the production and modulation of sound vibrations produced by the vocal chords. The vocal chords are vibrated by air forced out of the lungs by the diaphragm. As the vocal chords vibrate, the larynx muscle modulated the pitch by tightening or relaxing the vocal chords. Many of these structures can be seen in Figure 1 below.

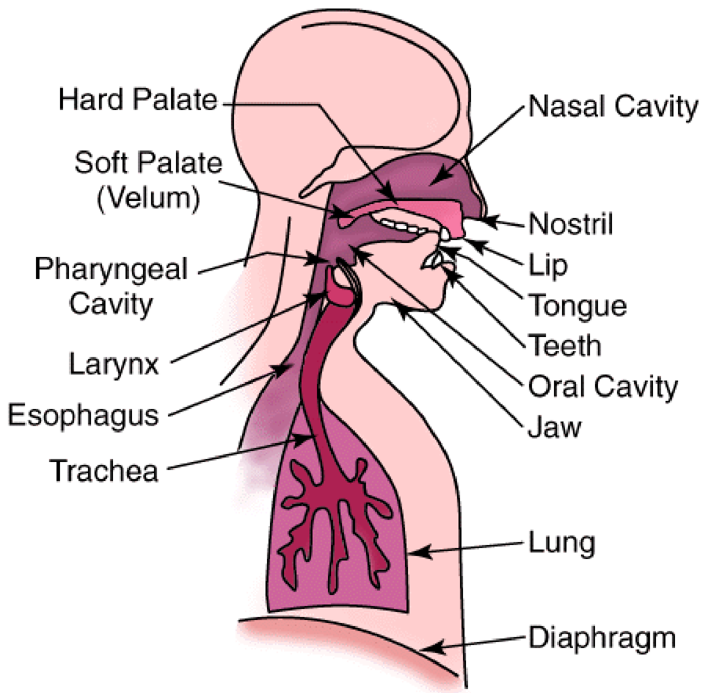

Figure 1: A diagram of the anatomical structures used in the production of normal speech in a healthy individual.

The fixed filters of the throat and nasal passage further modulate this sound. The sound is then modulated into language by the tongue, lips and facial muscles. The hard palate, soft palate and teeth are also important in formation of certain consonant sounds.

Many of the difficulties people have with speaking stem from the inability to produce vibrations across the vocal chords. This could be due to a laryngectomy, paralyzed diaphragm, cerebral palsy, as well as other various disorders. Our device seeks to replace the functions of the diaphragm, lungs, vocal chords and larynx by generating a column of pressurized vibrations that the facial muscles can modify. At least some function of the tongue, lips and facial muscles will be required for use. The hard and soft palates, also need to be intact in order to generate consonant sounds.

Design Requirements

One of the biggest complaints from patients with communicative disorders is that the devices out there to help them speak are slow and lack the ability to add emotion to what the user wants to express. This delay, between when the user thinks a phrase and when they are able to actually communicate, can make the user feel unintelligent or that they are being perceived as unintelligent. It can also leave them out of a conversation since they cannot produce language within the normal pause of a conversation. In addition, they are unable to add emphasis or inflection to what they want to say. These are the issues that our client, Dr. Lawrence Kaplan, has asked us to address. Every day he encounters patients that are frustrated with the means of communication to which they are limited. Many of his patients give up trying to speak and let others do it for them. Our client is looking to break away from the conventional communicative devices that are on the market today. He would like us to come up with something new that allows the user to have more spontaneity when they speak, as well as the ability to demonstrate emotions such as irritation or excitement when they communicate. By doing this, we can hopefully “bridge the gap” for people with communicative disorders and help them to communicate in a way that feels more natural and comfortable.

The main feature that our client is looking for in the device is that it be phonetics-based instead of text-based. With sounds at their disposal instead of words, people have a wider variety of things they can say. They can put more personalization into their speech and incorporate slang. Working from this idea, we hope to make the device fast and intuitive so that the user feels that they are better able to express themselves, as well as feeling more apart of conversation.

Previous Work

This project began two semesters ago with the establishment of the principle behind our current design. We used musical hardware in order to generate a “proof of premise” prototype in order to test the principal behind the idea. The hardware used included a Kaossilator as an input source and a talkbox as a mechanical sound source.

The premise of our idea is that sound can be funneled into the mouth and shaped into meaningful language, in the same manner as naturally-produced sound. By using the generated sound, the user does not need to produce his/her own sound, which is a problem for individuals with certain speaking difficulties. The device replicates the function of the diaphragm, lungs, and vocal chords. It generates the signal, much like the vocal chords, and the signal is pushed to the user’s mouth using the same principles as the diaphragm and lungs. One stipulation, however, is that the user must have control over his/her facial muscles so that he/she has the ability to shape the sound into understandable language.

The device works by taking an electronic input signal and converting it to mechanical sound. For our prototype (Figure 1), the input came from the Kaossilator, using a setting which produced digitized vowel sounds. This input was converted by a compression driver in the talkbox to mechanical sound and funneled into the mouth by a vinyl tube. This sound was then shaped into language as the facial muscles modulated the frequencies generated by the talkbox.

Last semester, the prototype was miniaturized using our own custom components (Figure 2). Our final design consisted of two parts: an amplification system and a compression driver. First, we input the human sound into the filter. This human sound was a pre-recorded sound with a continuous 'aa' syllable, played through an iPod. The sound source produced low amplitude signals which were attenuated when passed through the tube of the compression driver. Also, with a low voltage it was not possible to achieve resonant frequencies to occur at medium range frequencies. Hence we first tested the compression driver for different voltages and frequencies in order to produce a reasonable output. While testing we found that there is need of at least a gain of 15 to produce an audible sound.

In our search for an amplifier, we looked for something that would give us a reasonable gain, have a good heat sink, and would not produce too much noise. Professor Mark Allie of Electrical Engineering works with acoustics. When we consulted him regarding the amplifier he provided us with one which was quite compatible with our compression driver. It had a gain of 21 with a reasonable heat sink.

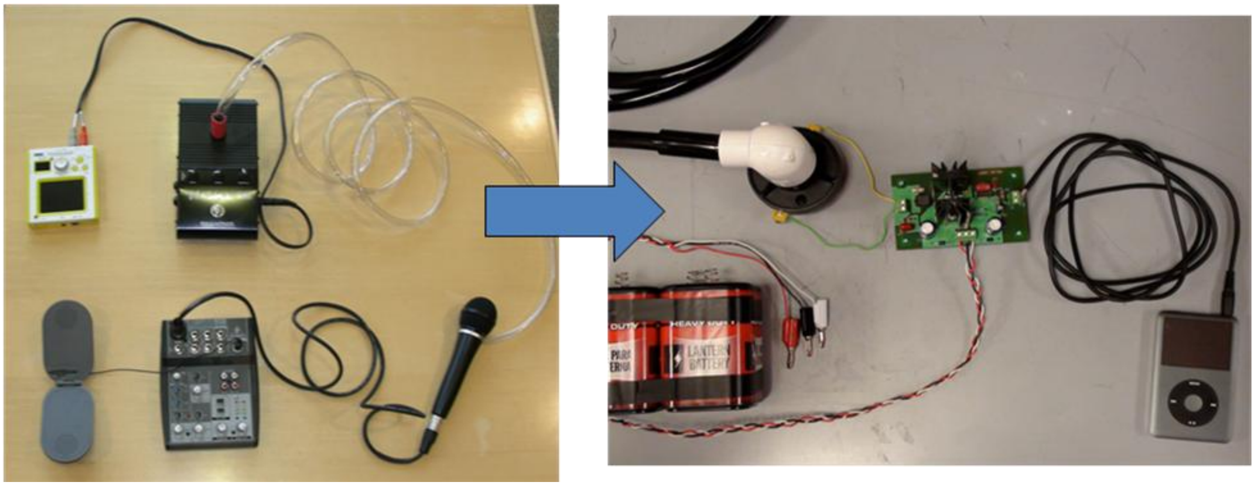

Figure 2. Our design two semesters ago was made up of musical parts. Last semester, our own custom components amplified the signal and pushed it through a tube into the user's mouth for modification.

After testing for different voltage and frequencies, the sound files were saved in an iPod, and this was then fed into the amplifier through headphones. The amplifier was powered using two 6 V batteries which provided a +/- 12 V power supply with ground connection. The set up took about 8 mA of current to run smoothly. The amplified signal was fed into the compression driver which sent high pressure sounds through the tube into the user's mouth. These sounds were then converted into words via lip sync. This concept is very similar to the way humans produce sound. In order to speak using this device, one only needs to use his/her facial muscles. The rest is done by our device.

Work also needs to be done on the compression driver. While our current compression driver is relatively small, it is necessary to go smaller in order to continue the miniaturization process. We also need to develop a method of securing the connection between the driver and tube. The loose fit right now allows some air to leak out, decreasing the pressure in the tube and ultimately decreasing the sound that reaches the user's mouth. A secure connection would maximize the signal that enters the tube. Finally, we need to settle on a sound source which will provide the most natural and realistic speech.

Testing of the device showed that the understandability of the speech was best when the tube was placed about a centimeter in the mouth. This way the teeth could be used to occlude the sounds and create consonant sounds. It was found that the tube diameter did not affect the sound quality for the testing that we performed. We determined that a small tube would be best for the future design.

Inflection and emotion were also important factors in proving the principal of our idea. We tested the ability of others to understand whether the expression was a statement, exclamation or question, It was determined that the average rate of understandability was 91.11%, on the second iteration of the expression. This result can be seen in Figure 3.

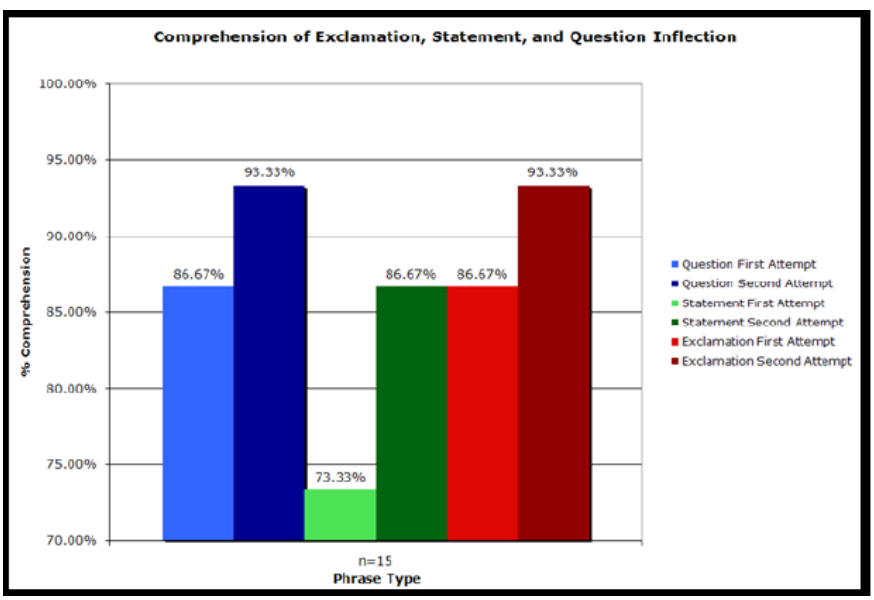

Figure 3. The inflection was understood 91.11% of the time on the second iteration of the expression.

The previous work on this project has laid the groundwork for the direction we will take this semester. We hope to integrate the hardware into a small, more compact device that is easy and comfortable for the user.

Final Design

Currently, our prototype (Figure 4) consists of a sound source that is fed into an amplifier. The amplified sound is fed into the compression driver and driven at high pressure through the tube into user's mouth. The high pressure sound enables the sound to be modified by the mouth acting as a filter.

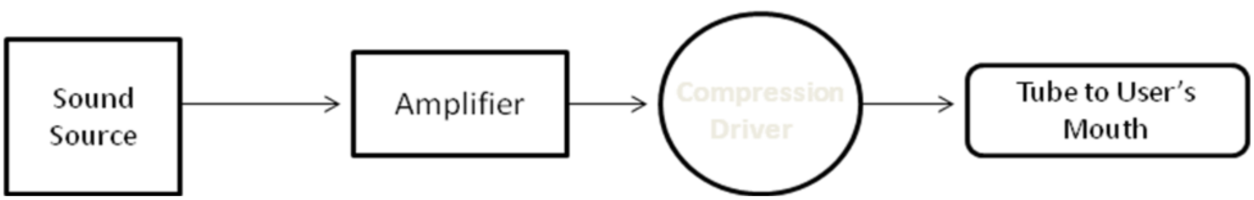

Figure 4. Block Diagram of Device

The sound source that is fed into amplifier should have a frequency of 800 HZ to 2 KHz in order to get recognizable speech at the output. The human speech is in the range 300 Hz to 3 KHz. Therefore, our tested data works well within the range.

The amplifier used for the device used a low voltage audio power amplifier chip, LM 386 (Figure 5). The chip has a very low quiescent drain current of 4 mA. The gain of the circuit is about 25. It provides an output current of up to 300 mA (AC) for the device to function effectively. At this current, the volume of the sound coming out of the compression driver is loud enough and need not be associated with a microphone. Two 0.1 μF capacitors have been used in the circuit in order to avoid fluctuations and instability in the circuit. A 110 µF capacitor has been used at the output to get rid of any DC offset. At the bypass pin 7, another 0.1 µF capacitor has been used to overcome power supply rejection at higher frequencies.

In order to power our amplifier, we have employed three AAA Enercell batteries used in series to provide a voltage of 4.5. It is rated at 1.2 V and 850 mAh. Considering our DC current consumption is 30 mA, our device should theoretically last for 28 hours of nonstop use. However, we have incorporated a switch with our device which would reduce the power consumption by a large amount. This also overcomes the need of a heat sink. Thus, the amplifier is very light, small and efficient.

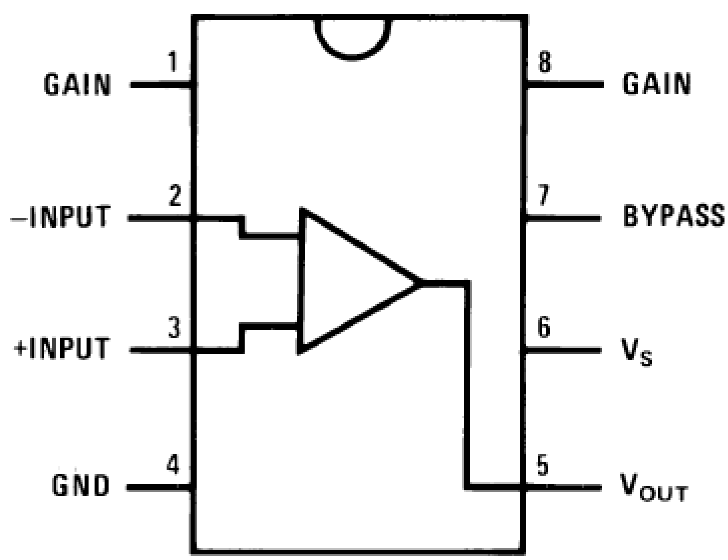

Fig 5: Pin diagram of LM386 low voltage audio power amplifier chip

Sound Source

A number of different sound sources are possible to interface with our hardware. The first we used is the Kaossilator. This is a dynamic-phrase synthesizer that is able to produce a wide range of frequencies and intonations when the touchpad is pressed. By using this device, the user is able to control exactly when sound is pushed through the tube with his/her finger. It also introduces the concept of inflection into the output voice. This allows the user to better express his/her feelings and also helps others understand what is being said.

Unfortunately, the Kaossilator produces a noticeably robotic sound. This is something that usually makes the disabled individual feel different from others and should therefore be avoided. To combat this, we worked on creating a sound source that mimicked a human voice as closely as possible. Our first attempt was using an iPod Touch to play recorded sound files of a constant male tone. When fed through the amplifier and horn driver, a much more realistic sounding voice was produced. The issue with this was the inability to quickly change the pitch or vowel sound which is often quite desirable. In addition, it was very difficult to stop the sound between words, which led to the slightly less recognizable speech.

Our next attempt at reproducing the human voice was the development of an application for Android-based phones. The entire Software Development Kit (SDK) is available online, and programs are written in Java. An application is more accessible to the general public, and the interface is more user-friendly. The application we developed has different vowel sounds available on the screen to play, along with a few consonant sounds for users who are unable to produce these naturally at the ends of words. The user simply presses the sound he/she wishes to be played, and this will be generated and pushed out of the tube. A few improvements need to be made to perfect the application, but it holds great promise for providing a simple, human-like sound source.

Packaging

The final device was packaged into a compact hip pack using readily available commercial parts. First, to connect the compression driver to the tube, a poly(dimethylsiloxane),PDMS, gasket was cast over the end of the threading on the driver. Approximately 50 grams of the polymer was mixed as 10 parts base to 1 part crosslinking agent as per manufacturing instructions. The compression driver was then held in place in the liquid PDMS by a foam mount. The entire apparatus was then heated at 65° C for 15 hours to cure the polymer. The solid mould was then cored out to the diameter of the tubing to ensure a tight seal. The remaining device was fabricated out of PVC plumbing fittings and a polycarbonate backing.

The top of the device is a 3" to 1 ½" reducing bushing, and the bottom segment is composed of a 3" diameter PVC connection fitting. The two segments are held together using bolted strips of cut aluminum, held with standard ¾"Phillips head bolts and nuts. The bottom cap of the device was cut out of a ¼" sheet of black poly carbonate. In the bottom cap there is a 9/16" hole cut for the output of the stereo cable to the input device. There is also a ¾" hole cored into the cap to hold the screw-fit on/off switch. The cap and bottom segment have 3 aligned, pre-drilled holes, so the cap can be screwed on with ¾" Philips head screws for easy access to the internal components.

The final aspect of the packaging is the top cap. In fabricating the sound-sealed gasket for the tubing connection, leftover cast PDMS was left on the periphery of the compression driver. This PDMS was then cut to the inside diameter of the reducing bushing and press fit into the narrow opening. From this the tubing can extend from the device to the user's mouth. The final aspects of the design is two hip attachment straps that also use segment attachment bolts as anchor points and will secure the device to the user's hip for portability. This simple packaging design allows for easy access to the electronics inside, including the batteries. When the device runs out of battery, the user must simply unscrew the three screws on the back end of the device and insert new batteries.

Problems Faced

While much of the work we have done this semester has given us encouraging results, certain problems still need to be addressed. We have already theoretically determined that the battery life should be sufficient for the user of our device. However, it is still necessary to confirm this fact. We also need to create a convenient method of recharging the device without taking it apart to replace the batteries.

In addition, we have spent some time this semester characterizing the sound output of the device. We have found that the user is able to recreate the formant frequencies seen in their natural voice using the device. However, understandability is still difficult with some of the more percussive consonant sounds. A way of making it easier to add percussive sounds to the speech would improve the understandability of the device.

Finally, in our meeting with WARF earlier this semester, we were told to try to find a physical aspect of the device the device that was novel and crucial to the functionality of the device. While the idea is certainly novel, at the beginning of the semester the physical components were not. We still need to identify exactly what about the device could be patented. We have created an amplifier ourselves that we hope to improve and present to WARF when the provisional patent deadline comes around this fall.

Future Work

This semester, the device has come a long way in the improvements to the components and overall sound quality. There are, however, some improvements that could be made in the future. We have characterized the sound output of the device using a Fast Fourier Transform[9-13] and found that the user is able to recreate the formant frequencies seen in their natural voice using the sound input from the device. There is still difficulty in understanding some of the more percussive consonant sounds, however. In the future, a more comprehensive application with vowel sounds and some hard consonant sounds would help augment the speech even more and improve the understandability of the device. Also, the input sound currently being used is a prerecorded human voice. While we have seen great success at mimicking the formant frequencies of natural speech using these prerecorded sounds, noise exits because of the amateur quality of the recoding. Access to a sound studio with better recording equipment and reduced background noise would also improve the understandability of the device.

In addition to improving the understandability of the device, we would like to improve convenience by replacing the 1.5 V batteries currently being used with a rechargeable battery that has a longer lifespan. Doing this will make using the device easier because one would not have to open it up to replace the batteries and, ideally, the battery life would be long enough for the user to speak with the device all day without it running out of power.

Finally, characterization of a unique aspect of the device was a goal for this semester. We did develop our own amplifier for use with the device that we will present to WARF. However, our testing on tube length was rather inconclusive. It showed that some lengths had greater amplitude of desired frequencies and lower noise, but there was no clear trend in the data. This result suggests that the results may have been affected by how well the user was using the device. Further testing on more subjects would give us better statistical data to suggest whether or not there is any significance to the results we saw this semester, or if in fact tubing length does not greatly affect the quality of speech produced by the device.

Limitations

Since the user is required to lip sync in order to effectively use this device, damaged facial muscles and nerves that are involved in speech would inhibit the user from using our device.

The orbicularis oris muscle which surrounds the lip area is the main muscle involved in speech. Any damage to that muscle would prevent the user from using the device. Damage to other muscles like the buccinator and zygomaticus major could also limit one from effectively using this device. However, some words could still be produced in spite of damage to these two muscles.

Out of the 12 cranial nerves that are present in a human, the hypoglossal nerve is involved in articulation of speech and innervates the muscles associated with it. Hence, any damage to it could also prevent the user from using the device.

Works Cited

[1]EMedicine. “Cerebral Palsy.” http://emedicine.medscape.com/article/1179555-overview

[2]National Institute of Neurological Disorders and Stroke. “Cerebral Palsy: Hope Through Resarch.” http://www.ninds.nihngov/disorders/cerebral_palsy/detail_cerebral_palsy.htm#88033104

[3]Gait Analysis Laboratory. “Cerebral Palsy Program/Guide.” http://gait.aidi.udel.edu/gaitlab/cpGuide.html

[4] National Cancer Institute. “Larnygeal Cancer Treatment.” http://www.nci.nih.gov/cancertopics/pdq/treatment/laryngeal/patient/

[5] Image: http://72.233.24.94/~worldcf/wp-content/uploads/2008/12/f3.jpg

[6] R. Patel. Phonatory control in adults with cerebral palsy and severe dysarthria. *Augmentative and Alternative Communication*; 2002, Vol. 18, No. 1, Pages 2-10.

[7] A Waller, F Dennis, J Brodie, A Y Cairns. “Evaluating the use of TalksBac, a predictive communication device for nonfluent adults with aphasia.” *International Journal of Language & Communication Disorders*; 1998, Vol. 33, No. 1, Pages 45-70.

[8] http://www.nidcd.nih.gov/health/statistics/vsl.asp Feb 15, 2010.

[9] Guze, S.B., Brown, O.L. Psychiatric Disease and Functional Dysphonia and Aphonia. Arch Otolaryngol. 1963. 76:1 84-87.

[10] Brigham, E. Fast Fourier Transform and Its Applications. Prentice Hall. 1988.

[11] Oppenheim, A.V., Schafer, R.W. Digital Signal Processing. Prentice Hall. 1975.

[12] Smith, S.W. The Scientist and Engineer’s Guide to Digital Signal Processing. California Technical Pub. 1997.

[13] Blumstein, S.E., Stevens, K.N. Acoustic invariance in speech production: Evidence from measurements of the spectral characteristics of stop consonants. J. Acoust Soc. Am. 66:4. 1001-1018.