Development of an Electrolarynx Capable of Supporting Tonal Distinctions in Mandarin

Bicky Shakya '14, Vishal Bharam '14, Alexander Merchen '14
Faculty Adviser: Prof. David J. Ahlgren
External Advisers: Dr. James Heaton (Massachusetts General Hospital), Mark Robertson (Griffin Laboratories, California)

ABSTRACT

Tone distinction is a vital part of languages such as Mandarin, where the same words said in different tones could mean vastly different things. For people who have lost their voice box, usually due to the cancer of the larynx, making this tone distinction becomes impossible with currently available Electrolarynxes, which are limited to monotones. Further, due to a lack of tone distinction, the resulting voice sounds 'robotic' and unnatural. Wan et. al recently developed an electrolarynx (EL) capable of tonal control using a trackball, and TruTone® from Griffin Laboratories is another EL that features tone modulation via capacitive buttons. These designs feature user-interfaces which make it hard for pitch-changing during speaking, especially for tonal languages. This senior design project focuses on developing an EL-based system capable of modulating vocal frequencies so that tonal distinctions can be made for Mandarin. We have developed a novel user interface using a gyroscope to let user select and produce the tones necessary to converse in Mandarin. The gyroscope is embedded into a custom-made EL body along with a microcontroller (ATMEGA328) and additional circuitry for digital-to-analog conversion, voltage-controlled oscillation and amplification. Our system is capable of accurately reproducing the four tones in Mandarin, along with the fifth 'non-tone' as well as taking care of tone transition rules in Mandarin. Initial results have shown as much as 90% accuracy in the identification of the four individual tones produced by our EL system and understandable speech with multiple syllables.

BACKGROUND

Pitch is normally created when air pressure from the lungs vibrates vocal folds of the larynx. The vibrations of vocal folds along with movement of the lips and tongue is used to generate human voice. When people undergo laryngectomies (larynx removal), the affected individual cannot speak. Every year 3,000 Americans (1 in 100,000 globally) undergo laryngectomies due to laryngeal cancer. There are only a handful of techniques available for the patients to regain their voice such as Silent Speech Interface and oesophageal speech, which are either impractical for patients or difficult to master. The most popular device to use is known as an electrolarynx (EL), which is a handheld device that creates a supplementary tone that the patient can use to talk. For tonal languages, such as Mandarin, single-tone ELs will not allow a native speaker to distinguish between two similar words. Various attempts have been made to create an EL that can create the tones necessary in Mandarin, such as a trackball interface that allows user to vary their tones (Wan et. al). Nonetheless, most of these approaches are still in research phase and do not offer a user-friendly interface.
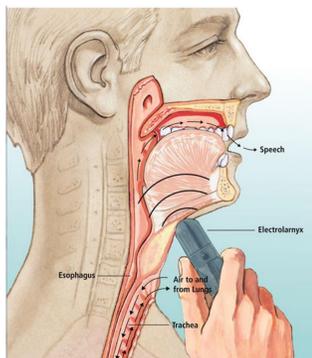

Fig. Using an Electrolarynx

http://store.inhealth.com/category_s/60.html

Fig. TruTone® From Griffin Laboratories

http://www.griffinlab.com

USER INTERFACE

In order to give electrolarynx patients the ability to voice tonal distinctions necessary for Mandarin, an embedded electrolarynx system was used to let the user choose the tones he or she would want to voice so that conversational Mandarin would be possible. A gyroscope-based user interface was implemented inside the body of the EL system. The user would hold the system against his neck at a particular position in order to calibrate the system and then move the electrolarynx body around in four distinct axes in order to activate the four different tones of Mandarin.

TECHNICAL APPROACH

For creating the four tones of mandarin, we used a lookup table-based approach. Audio files of actual people (female) voicing the four tones were obtained. These audio files were then put through an open-source phonetic analysis software called Praat. Praat allowed us to obtain the pitch contour of each of the four tones i.e. a plot of frequencies versus time. The audio files were then sampled at every 0.001 seconds to obtain the frequency values (in Hz). The frequency values were then exported from Praat and put through a custom MATLAB script for processing. After processing, we had four distinct lookup tables i.e. a table of frequency values for each of the four tones. These values were then loaded to the flash memory of our microcontroller .The microcontroller then took these frequency values and converted them to digital signals for further processing.

- **Microcontroller**

In order to process the lookup tables based on the frequency profiles of each of the Mandarin tone, a real-time controller is required so that when the user chooses a tone, it can be picked from the controller memory and converted to a digital output signal. After reviewing options we decided to choose the Atmel ATmega328 based Arduino Nano due to its: Small size (0.7" X 1.7"), and its compatibility with other components.

- **Voltage Controlled Oscillator (VCO)**

The digital signal from lookup tables is converted to an appropriate analog signal, which is used as an input to drive voltage controlled oscillator (VCO) that generates waveforms of desired frequencies at certain voltages and duty cycles. The VCO in our design is driven by the LM291 operational amplifier, a pair of NPN transistors.

- **Digital to Analog Convertor (DAC)**

A DAC is used to convert the digital signal generated using the lookup tables to analog signal that controls the VCO. The Adafruit 12-bit D2A chipset (MCP4725) was used because of its ability to get a rail-to-rail conversion. The final output of the DAC can vary voltage from 0 V to 5 V, with 12 bit resolution.
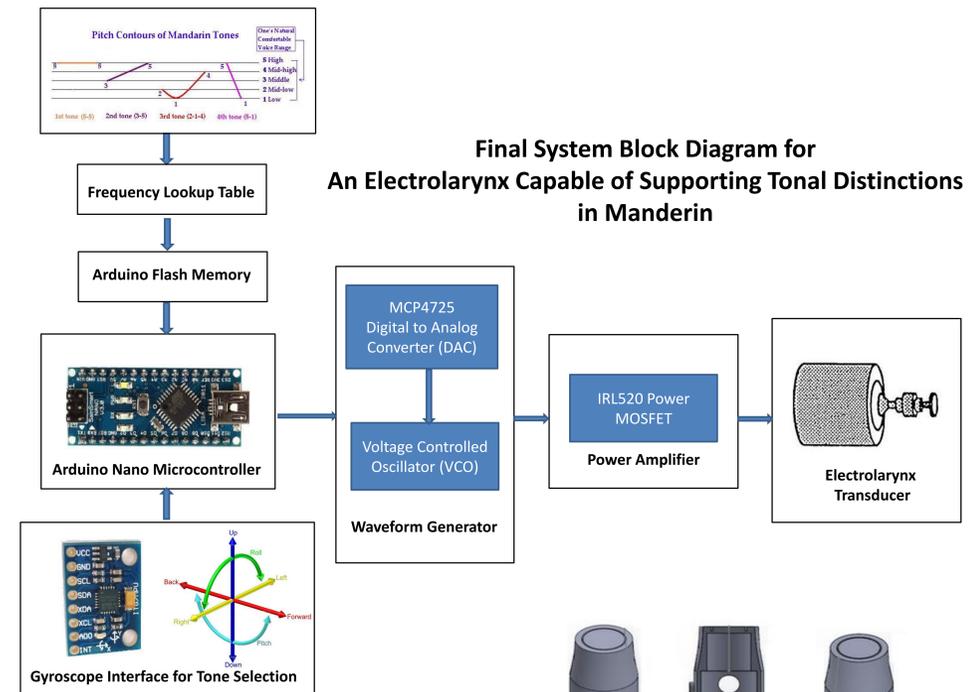
- **Power Amplifier**

The VCO-generated output frequency signal was fed into a power amplifier (IRL520) which convert these low-power driving signals into a strong enough signal to drive the transducer from the EL.

- **Transducer**

The Electrolarynx transducer is the mechanical device on the EL that produces the vibrations needed to produce the actual speech. The transducer consists of a magnet assembly (coil) that vibrates as a function of the input signal. This in turn vibrates a coupler disk, which is the head of the EL, the part that a patient actually places against their throat to phonate.

- **PCB**

The final verified system circuitry was built as a stand-alone embedded EL system on a printed circuit board (PCB) that consists of VCO, DAC, power amplifier and Arduino Nano.

Final System Block Diagram for
An Electrolarynx Capable of Supporting Tonal Distinctions in Manderin

Gyroscope Interface for Tone Selection

ADDITIONAL FEATURES

- Volume Control
- Tone Duration Control
- Can follow Mandarin Tone Transition Rules
  - Half tones
  - 5th No Tone

Fig. Final Mandarin EL Body

RESULTS

After the complete prototype was set up, a set of listening tests were conducted in order to verify the understandability of the four tones of our Mandarin-EL system. The tests were conducted on 7 subjects: 6 native speakers and 1 non-native speaker. An accuracy rate of ~90% was recorded from tests involving the voicing of the four Mandarin tones. Listeners reported clear understanding of the four individual tones as well as groups of syllables. However, certain flaws were inherently present. When voicing words such as ba and pa, the mechanical nature of the sound produced from the EL caused slight confusion in understanding. Nonetheless, participants were still able to identify syllables based on context. Also, the gyroscope interface needed practice, especially while voicing long sentences. Lastly, although the Mandarin EL tones were generally understandable, the style and voicing of Mandarin varies from person to person and region to region, leading to slight confusion.

REFERENCES

Maddox et. al, *Trends in Total Laryngectomy in the Era of Organ Preservation: A Population Based Study*, American Academy of Otolaryngology (Head and Neck Surgery Foundation), 2012

Wan Congying, Wang Erqiang, Wu Liang, Wang Supin, Wan Mingxi. *Design and evaluation of an electrolarynx with tonal distinction for Mandarin,* IEEE Audio, Language and Image Processing. 64(6), March 5, 2013, 51(2): 291-96

Mandarin Tone Profile Image from http://web.mit.edu/~jinzhang/www/pinyin/tones/#single