MYOVOX: A PLUG AND PLAY DEVICE EMULATING A MOUSE AND KEYBOARD USING SPEECH AND MUSCLE INPUTS

Authors: Matt Belgiovine, Mike DeLiso, and Steve McGill
Advised by Dr. Daniel D. Lee

The University of Pennsylvania

ABSTRACT

We are creating a portable system enabling people to interact with a computer through select speech and muscle activity inputs. We will use an embedded system capable of taking these inputs and interface with a computer emulating a standard plug-and-play keyboard and mouse. Sphinx open-source programming will be the principal speech recognition technology while surface electromyography (sEMG) will be the primary muscle recognition interface. The system created is targeted toward programmers suffering from injury that negatively affects their ability to utilize a traditional computer interface system. To this end, there are four specific goals of the project:

- To create a portable, speaker dependent speech recognition system based off of the Sphinx speech platform.
- To enable speech recognition word banks targeted to the end user that includes the military alphabet, keyboard commands and information technology (IT) professional specific libraries.
- To enable mouse-emulating x and y axis computer interface through the use of muscle activity inputs.
- To interface with any computer's Universal Serial Bus (USB) port and push American Standard Code for Information Exchange (ASCII) data to control any operating system including BIOS, Windows, Linux and Mac systems. .

1. INTRODUCTION

The ability to work at multiple workstations is a necessity in today's work environment. In order for an IT professional to maximize their value to any given organization, they must be able to complete their work at different workstations. This lack of mobility is particularly evident for those IT professionals unable to use a traditional keyboard and mouse. Speech recognition is an often used solution for these individuals, but the current systems restrict the end user to one computer. IT professionals affected by disease or trauma are currently limited in the value they can provide to firms by the lack of speech recognition products designed for their needs.

A solution to this problem is coding the speech interface software onto a pocket-sized embedded computer. The embedded system will be small and portable, capable of being implemented on separate computers without much reconfiguration. The device will communicate with other computer devices through USB acting no differently than a plug and play mouse and keyboard. The embedded system will receive speech inputs from a microphone headset.

Speech recognition technology has progressed significantly over the past decade. Carnegie Mellon University (CMU) has developed a series of open-source speech recognition systems. MyoVox will implement PocketSphinx, a real-time continuous speech system offered by CMU. Programs encoded on the MyoVox embedded computer will operate PocketSphinx to recognize the speech inputs relayed by the microphone.

Specific muscle activity inputs will also be recognized by the MyoVox system, enhancing ease of use and functionality for its users. Using the movement of any muscles, MyoVox will supplant traditional mouse usage. Actuators designed to measure movements in two axes will enable the end user enhanced control over a computer.

By using speech recognition of the military alphabet, keyboard commands, end-user specific word libraries, and muscle activity recognition to control graphical interfaces, MyoVox will successfully emulate a keyboard and mouse system.

2. STRATEGIC PLAN

The MyoVox project will be designed with the end user's needs as the driving specification. With this in mind, the approach will follow a route that solves the problems of current systems. The hardware and software requirements will need to satisfy the user's needs and, as a secondary concern, be as cost-efficient as possible.

2.1. Theory of Operation

The concept behind the device is that a user will plug the MyoVox device to their computer. Instantly, the computer will recognize the MyoVox device as a keyboard. The user will then attach the MyoVox muscle sEMG sensor to a described position on their body, and plug a microphone into the MyoVox device.

As the user talks, their commands will be recognized by the MyoVox device and will be accompanied by the appropriate behavior on their computer. For instance, the user may utter "Control Charlie." Note the use of the military alphabet to ensure accuracy ("Charlie" denotes the letter "c"). This keyboard sequence (ctrl+c) will then be sent to the computer. A word bank will be compiled for valid commands. This word bank will include the military alphabet, keyboard commands, and be extended for certain UNIX or programming environments (commands like 'cat', 'if', 'else', etc.). There will also be commands for moving the mouse, such as "move mouse 10 pixels left."

As the user moves the portion of their body that is attached to the MyoVox muscle pack, the computer's mouse will move to reflect the user's motion. The MyoVox muscle pack will include a battery that can be recharged by the user.

We are advancing our system design in a three pronged approach. We are analyzing the inputs to our system, the outputs of our system, and the underlying black box that maps inputs to outputs. Our input system needs to be modular so that multiple forms of user input signals can be handled. This is important for future revisions of our hardware, where more sophisticated forms of human interface can potentially be devised. This is a form of future-proofing. We need our output system to be just as modular as our input system so that the MyoVox device can be extensible to the many interfaces that IT professionals may encounter. The processing subsystem will be designed to take advantage of existing software packages that can help meet user demand, but have not been connected together. The important part of this subsystem will be the processing of each input module and the formatting of this processed data for use by the output modules.

While we underscore the need for modularity in the input and output system, we will ultimately need to narrow our scope for an initial set of inputs and outputs, rather than a vague understanding of the I/O infrastructure. In our design decisions, we will identify the most common interfaces that end users encounter in their work.

In the underlying processing engine, there will be three program series. The initial program will take the input and

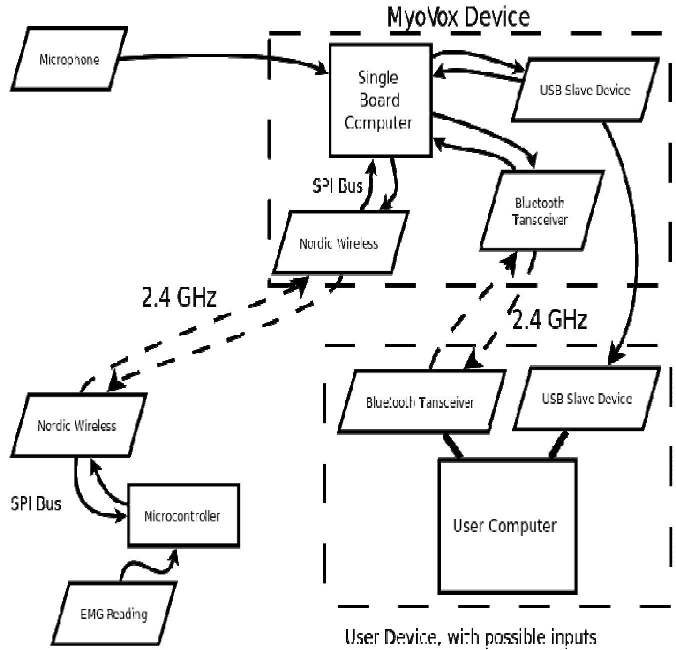

Figure 1: MyoVox System Diagram

map it to a command packet. For instance the voice module will be an InputProcessor. An InputProcessor (e.g. VoxSpeech in Figure 4.1) is a program that constantly samples the input signal, processes that input, and sends a packet to the PacketManager program, a Human Interface Device (HID). The PacketManager program will take packets of the form 'ascsii|a' or 'word|cat', process these packets, and output the data in the standard data structure to the proper OutputProcessor. An OutputProcessor will buffer the data from the PacketManager through the USB and output the protocol formatted data to its specific output device.

2.2. System Specification

There will be two inputs that we will use for our MyoVox system. The first will be audio gathered from a microphone that is connected to the MyoVox computer engine via a 1/8" audio jack. The MyoVox system will be able to process user speech through this device in real time before outputting an HID command to the user's host computer. For our project, we will define "real-time" to mean that, within 1 second of uttering a command, the user will see the correct response from the MyoVox device. The MyoVox device should be able to achieve an accuracy of above 95%. We determined this requirement based on the performance characteristics of Dragon Naturally Speaking 10 which are cited to be 98%. The second will be muscle motion gathered from EMG data. This motion data will be sent wirelessly to the MyoVox device.

The MyoVox device will feature two output connections. A USB output connector will feed the user computer with relevant keyboard and mouse data without the user having to install device driver software in addition to that of standard mice and keyboards. The device driver will emulate that of the 104 key keyboard.

In addition to a USB connection, the MyoVox device will feature a Bluetooth keyboard interface. When the bluetooth feature is enabled, the MyoVox device will pair with a Bluetooth enabled device, and send Bluetooth standard HID keyboard data. The user's device will need to support the standard Bluetooth keyboard protocol. The Blackberry smartphone is an example of this device.

Based on user demand, the device, optimally, will be powered by the user's USB connection, or via a built-in rechargeable battery. However, the MyoVox device will be powered by an external power adapter due to uncertainty in the hardware that can meet performance demands of the system and the time constraints on the project. Time and hardware permitting, a battery system will be implemented.

2.3. Hardware Requirements and Design Approach

The audio input system needs to be able to deliver at least 16 bit audio, as this is the assumed case in the PocketSphinx codebase Also, the sampling rate should be at least 8kHz because PocketSphinx's Acoustic Model Libraries were created at this resolution and higher. The audio portion will be directly connected to the onboard computer.

The muscle input system will be implementing a surface EMG capturing and processing system to replace the simple flex sensor model. We will make use of the MA-411 EMG Preamplifier on every muscle that we are recording electrical activity from. The analog signal from this EMG preamp will go to the microcontroller for analog to digital processing. The signal will then undergo filtering and will be passed through the Teager Kaiser energy operator in order to extract maximum information from the signal. From here, amplitude and threshold based processing will be used to determine the intent of the user and the output so decided.

The muscle input system will relay data to the MyoVox device over a wireless frequency that can communicate at least 1 meter from the user's computer. We will use the Nordic chipset. This is because the team is most familiar with the chipset, and it has been acquired from Professor Fiene of the Mechanical Engineering (MEAM) Department for $15 per board, with a $15 microcontroller that interfaces with it. The microcontroller platform is the MAEVARM, a microcontroller developed by Professor Jonathan Fiene of the MEAM department at the University of Pennsylvania. The board runs an Atmega32U4 from Atmel with an 8MHz clock speed. The microcontroller requires less than 50mW of power, and can use far less depending on the pins that are used. The microcontroller will also serve as the analog to digital converter for the EMG readings. It samples from 3.8kHz to 153.8kHz, which is well above the required sampling rate necessary for EMG readings.

The MyoVox device will require a significant processing core in order to decipher speech using PocketSphinx. To satisfy the demands of the device, we need to investigate the single board computer market. The computers need to offer the capability for USB slave mode, include a Linux kernel, and offer the ability to process SPI for wireless communications with the muscle pack.

We will use the BeagleBoard for the single board computer. This offers users of the device a smaller initial cost, with the downsides being the additional physical size and less available RAM than alternatives. The smaller amount of RAM will affect the board in terms of processing speed for large datasets, but not in terms of raw processing power, which is dictated by the CPU. If PocketSphinx is bottlenecked by the deficiency of RAM, we may need to acquire additional RAM. The BeagleBoard's Package-On-Package RAM system enables this option.

2.4. Software Requirements and Design Approach

Our software needs to be robust at determining the word bank that we provide it. It also needs to be able to robustly work with interprocess communications. The software needs to be able to seamlessly produce data streams that conform to the communications protocols that we are implementing – i.e. it needs to follow the Human Interface Device (HID) protocol for USB and Bluetooth.

The reason for adhering to the HID protocol is that HID is standard across all user devices that will interface BIOS, Linux, Windows, etc. In order to program the device to communicate the HID protocol, we will be using a Linux operating system running on the MyoVox device that implements the USB Gadget API.

The protocol that we use must conform to the USB On-the-Go standard as this is what our hardware will use to communicate with the computer.

The software package that we will use for this project is the PocketSphinx project which runs on the ARM processing core (TI OMAP) that we will use for the hardware. The software has the ability to process speech in real time. It also can achieve word error rates under 10%.

3. DEMONSTRATION

Once the initial muscle recognition system is developed and constructed, the testing will require verification of the effectiveness of the sEMG data, the ability of the microprocessor to recognize and process the analog signals, and the viability of the wireless transmission system. We will require a multimeter for the test of the actuators and a Bluetooth capable transceiver.

For the speech recognition, we will first examine the defined language dictionaries through the testing programs provided by PocketSphinx. The testing programs provide an estimation of the word-error rate (WER) and sentence-error rate (SER) of the dictionary. Once the dictionaries have been finalized, we will test the real-time WER and SER by using a portable microphone that attaches to our laptops.

Once we are satisfied with the effectiveness of the program, we will implement it on to the BeagleBoard. Since this will be interfaced through both USB and Bluetooth, we will have two separate demonstration modules. In order to test both systems, we will be using our Linux machines for the USB test and a Bluetooth-compatible Blackberry 8310.

A test of the demonstration will be considered with the successful interpretation of speech data at or below a WER of 5% at a lag of less than one second. The muscular recognition system will be judged by its ability to recognize inputs at real-time pace.

4. FINANCIAL INFORMATION

In a project of this magnitude, there will be certain costs involved. Some of the hardware involved in the creation of MyoVox is not everyday material easily obtainable through the available university resources. Those material resources that are available often cannot be dedicated solely to MyoVox use.

While there are funds available to us the University of Pennsylvania for the development of our project, the costs we expect to incur exceed the $75 per person allowance. Therefore, we have applied to a number of external grant sources for further funding. As of now, we have applied for the IEEE Mini-Grant, CURF Vagelos Undergraduate Grant, and the Weiss Tech House Innovation Fund backing.

We are pleased to have been awarded the IEEE Mini-Grant worth $500. This will cover a substantial amount of our costs, but to fully realize the value in our project we hope to obtain more funding.

5. REFERENCES

[1] Ben Mosbah, B. "Speech Recognition for Disabilities People." Information and Communication Technologies, 2006. ICTTA '06. 2nd .Volume 1, 0-0 0Digital Object Identifier 0.1109/ICTTA.2006.1684487 Page(s):864 – 869.

[2] Davies , K.H., Biddulph, R. and Balashek, S. (1952) *Automatic Speech Recognition of Spoken Digits*, J. Acoust. Soc. Am. **24**(6) pp.637 – 642.

[3] D. Huggins-Daines, M. Kumar, A. Chan, A. Black, M. Ravishankar, and A. Rudnicky, "PocketSphinx: A free, real-time continuous speech recognition system for hand-held devices," in Proceedings of ICASSP 2006, Toulouse, France, 2006

[4] D. Huggins-Daines, and A. I. Rudnicky. "Combining Mixture Weight Pruning and Quantization for Small-Footprint Speech Recognition." in Proceedings of ICASSP 2009. 4189-4192.

[5] Paulson, L.D. "Speech Recognition Moves from Software to Hardware." Computer. Volume 39, Issue 11, Nov. 2006 Digital Object Identifier 10.1109/MC.2006.401Page(s):15 – 18.

[6] Roe, D.B. and J.G. Wilpon. "Whither speech recognition: the next 25 years." Communications Magazine, IEEE. Volume 31, Issue 11, Nov. 1993 Digital Object Identifier 10.1109/35.256880. Page(s):54 - 62.

[7] "Teager–Kaiser Energy Operation of Surface EMG Improves Muscle Activity Onset Detection."Annals of Biomedical Engineering. Springer Netherlands. 0090-6964 (Print) 1573-9686 (Online). Volume 35, Number 9 / September, 2007

[8] Woodland, P. "Speech recognition." Speech and Language Engineering - State of the Art (Ref. No. 1998/499), IEE Colloquium on. 19 Nov. 1998 Pages:2/1 - 2/5

.