

Cursor Control Using Voice and Facial EMG Signals

Grant Connell

I. Introduction

There are many applications for sound card EEG units that interact with humans for BCI and other control applications. Some of these applications allow disabled individuals to have improved communication with their caretakers and also allow them to use computers independently. Applications that only use head muscles or signals would be especially interesting and useful for individuals with extreme physical immobility. This paper describes a project that uses voice and facial EMG signals for that purpose.

The recent revision of the AM sound card EEG units allow for sample rates between 256 and 4096 Hz. The higher sample rates allow the units to be used for applications that use EMG signals in addition to EEG signals. This project demonstrates an EMG application that uses EMG facial muscles and sub-vocal EMG signals to control the cursor on the computer screen. The advantage of using EMG signals is that they are unaffected by external noise in the environment. In addition to controlling the cursor, the facial EMG signals can also be encoded to provide for the generation of the full alphabet and numbers. Voice EMG could offer the advantage of greater communication capabilities. However, the disadvantage, for voice input, is that due to a limited number of sensors only a subset of voice commands may be usable. Speech is generated using many different throat and facial muscles. The sensors pick up only the EMG signals from a subset of these muscles.

The goal of this project was to provide real-time control of the cursor on the screen using EMG signals. The project first approached using voice EMG signals but the recognition rates that were achieved with simple real time DSP algorithms was about 45% to 50%. The weakness in the processing was inconsistent real time feature generation. These recognition rates were too low to prevent a user from being frustrated in controlling the cursor. Using EMG facial signals a 90+% success rate was achieved and higher rates are possible through additional practice. The limitation is the repeatability of the user not the recognition software. In the end, the success rate has to be only high enough for a subject to want to use the software for cursor control. The facial EMG software used Morse code encoding to move the cursor. The voice software focused on a subset of commands such as stop, up, down, left, right, click, and double click.

II. Setup

The hardware was a dual channel AM sound card EEG unit coupled into the microphone input of a 3.2 GHz laptop computer running WindowsXP, SP2. For the voice configuration there were three AgCl electrodes (channel 1 +, channel 2 +, and common (-) of the amplifiers). In both configurations the DRL output of the EEG unit was connected to a wrist strap.

For voice EMG input, the AgCl electrodes were placed under the chin and on the right throat locations. Originally, two elastic bands were made to hold the electrodes in place. One band was placed around the throat while the other band was placed around the head/chin area. The electrodes were held in place onto the bands with Velcro. Although two channels of electrodes were used, only a single channel was processed in this effort.

After much testing of a given command set with the two electrode bands and achieving mixed testing results, the electrodes were moved to only under the chin (voice) and jaw (facial) locations. The common electrode (minus side of the amplifier inputs) was along the chin center with channel 1 plus (+) electrode

placed on the right side about 1 inch distance. This simplified the electrode placement with only one chin/head strap being used.

For the facial EMG configuration only two electrodes (for one channel) were placed on either side of the jaw. A second channel was used for reference to compare feature generation with electrodes placed only one side. A single headband (vertically oriented) held all three electrodes. Both channels of EMG data were used (added) to smooth out the noisy nature of the signal for feature processing. The electrode positions are shown in figures 1 and 2.

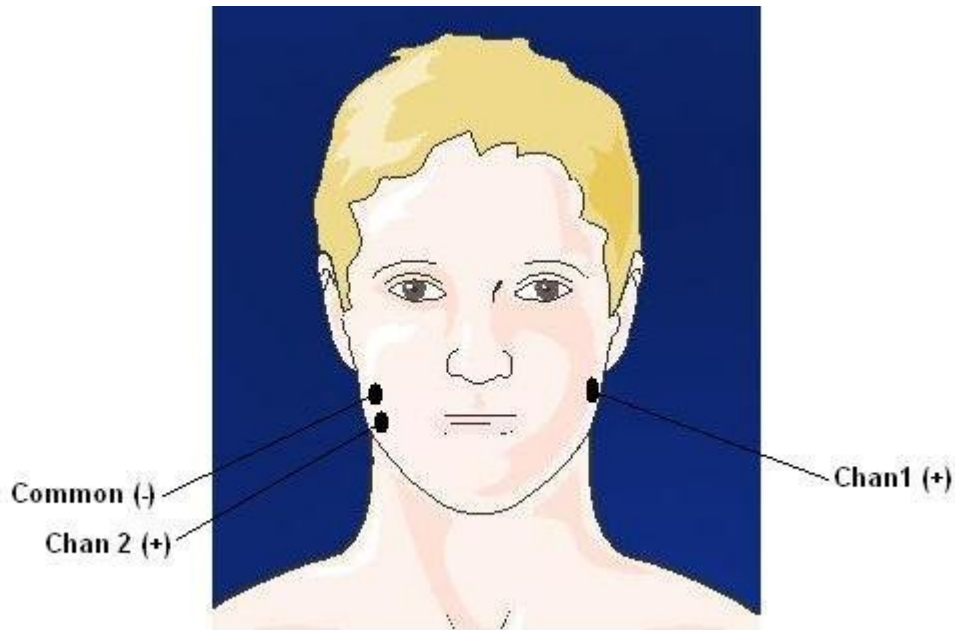


Figure 1, Facial Electrode Configuration

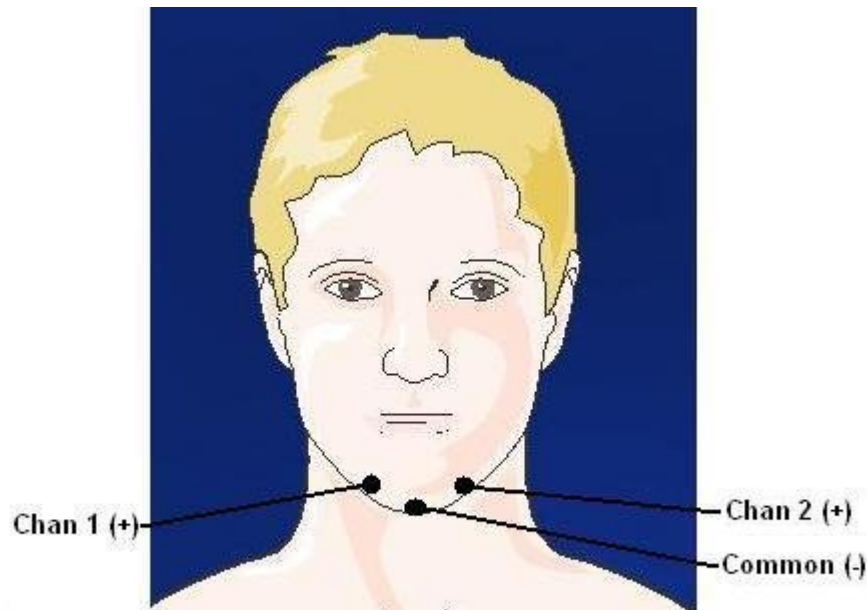


Figure2, Voice Input Electrode Configuration

III. Design

The system was run at a 1024 Hz sample rate with each channel being band limited from 35 to 450 Hz. The 35 Hz lower frequency limit was set to remove the Heart ECG signal. The sound data buffer was 2756 samples in length, captured 16 times a second. A notch filter was used to reduce the 60 Hz interference. The signals were converted to a RMS full wave rectification and then low pass filtered at 10 Hz to smooth the output waveform before being processed.

The software used was the NeuroProbe application developed specifically for sound card EEG units. A "Voice Cursor" element was added as an output device with two input channels. A design template was created to pre-process the data before being sent to the Voice Cursor element. The Voice Cursor element has menu options to bring up two additional screens, the cursor control screen and the vocabulary template input screen. The design is shown in figure 3.

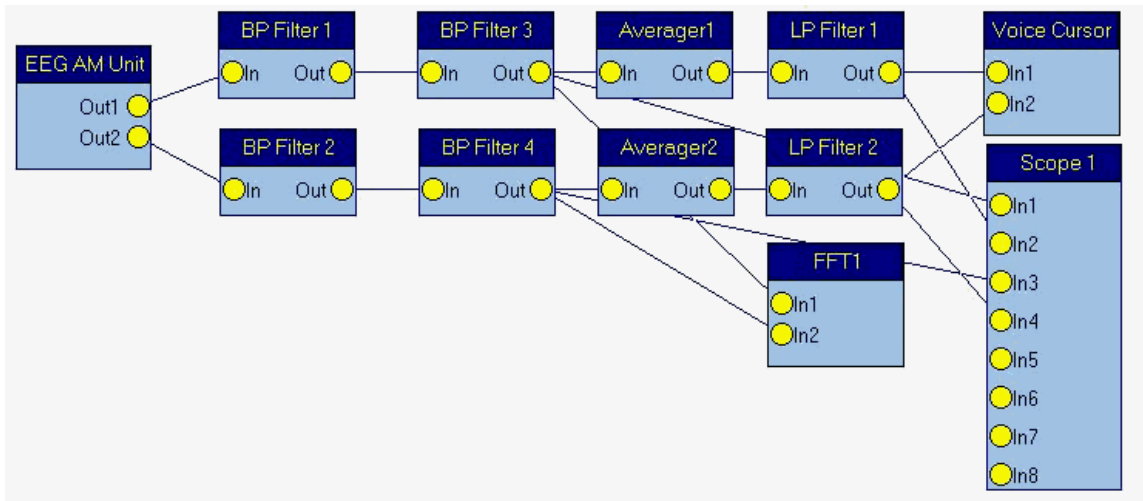


Figure 3, Voice and Facial EMG Design



Figure 4, Voice Cursor Control Screen

Element filter design summary:

1. EEG AM Unit Gain 0, both channels
2. Channel 1 frequency = 8192 Hz, Channel 2 frequency = 4096 Hz
3. BP Filter 1; Notch function, LP frequency = 58.5 Hz, HP frequency = 61.4 Hz
4. BP Filter 2; Notch function, LP frequency = 58.5 Hz, HP frequency = 61.4 Hz

5. BP Filter 3; BP function, LP frequency = 500.0 Hz, HP frequency = 35.0 Hz
6. BP Filter 4; BP function, LP frequency = 500.0 Hz, HP frequency = 35.0 Hz
7. LP Filter 1; LP frequency = 10 Hz
8. LP Filter 2; LP frequency = 10 Hz

IV. Pattern Recognition

For both EMG input configurations, a heuristic approach was used for pattern recognition (waveform matching) and to improve the probability of successful interpretation of the waveform. The advantage of the heuristic approach is that the template containing the matching waveform is tuned to the user. There is also the advantage that, in the case where a user has a degenerative disease, the vocabulary can be retrained to the changing speech characteristics. The first step was to establish a method to capture the waveforms associated with a command. The input was voice activated with the command and epoch being user selectable.

For facial input the input was compared to a threshold. If the signal was above the threshold a the input value was set at a constant value (200). Otherwise the signal was set to zero. The template waveform training used a similar approach in that the averaged waveform was compared to a 50% of a maximum constant. If the averaged waveform was greater than the threshold the template value for that point was set to the constant (again 200). Otherwise it was set to zero. For voice input, the epochs were averaged and scaled to a constant amplitude (90% of screen height) to compensate for different speech intensity levels. The waveform shape stayed the same (mostly) during speech intensity level deviations of ($\pm 25\%$). This made the recognition process less sensitive to signal amplitude variations.

The digital signal processing (DSP) of the waveforms used several common techniques for processing signals in noise. An average noise floor was computed to provide a base threshold level for voice signal detection. This average was frozen during the time the signal was above threshold. The following process steps led to the recognition of the waveform associated with a command:

1. The signal processing was started if the waveform exceeded a given threshold and continued until 1024 data points were gathered. This provided for 1 second of data. It implies that commands cannot be greater than 1 second in length.
2. The average noise floor was then subtracted from the waveform leaving only the active signal. The signal was also amplified to provide a greater numerical signal value for processing and display.
3. The signal then was normalized in amplitude to a constant. The peak of the waveform was compared to the constant (90% of the screen height) and the waveform scaled until the peak of the waveform was equal to the constant. This waveform was used for both command recognition and template generation.
4. The waveform was compared point-to-point to each template waveform. The absolute value of the difference (distance vector of each point) was summed successively across all of the compared

points. Some waveform templates have different lengths. The results were stored in a distance array containing the value of total distance for each template.

5. The distance array was then normalized for the different template lengths by increasing the distance value by the ratio of $1024/(\text{template length})$. This would compensate for some commands being longer or shorter. This is in effect a time warping of the command waveform.
6. The array index of the template with the lowest distance value was used as the command index for moving the cursor or clicking the mouse.

V. Testing

The voice and facial EMG designs included both a FFT and Scope element to observe the waveforms during testing. The FFT screen was useful for monitoring the 60 Hz component along with the total signal spectrum. The Scope screen displays the band-limited raw speech and low pass filtered waveforms. Figure 5 shows the test environment during speech capture and recognition.

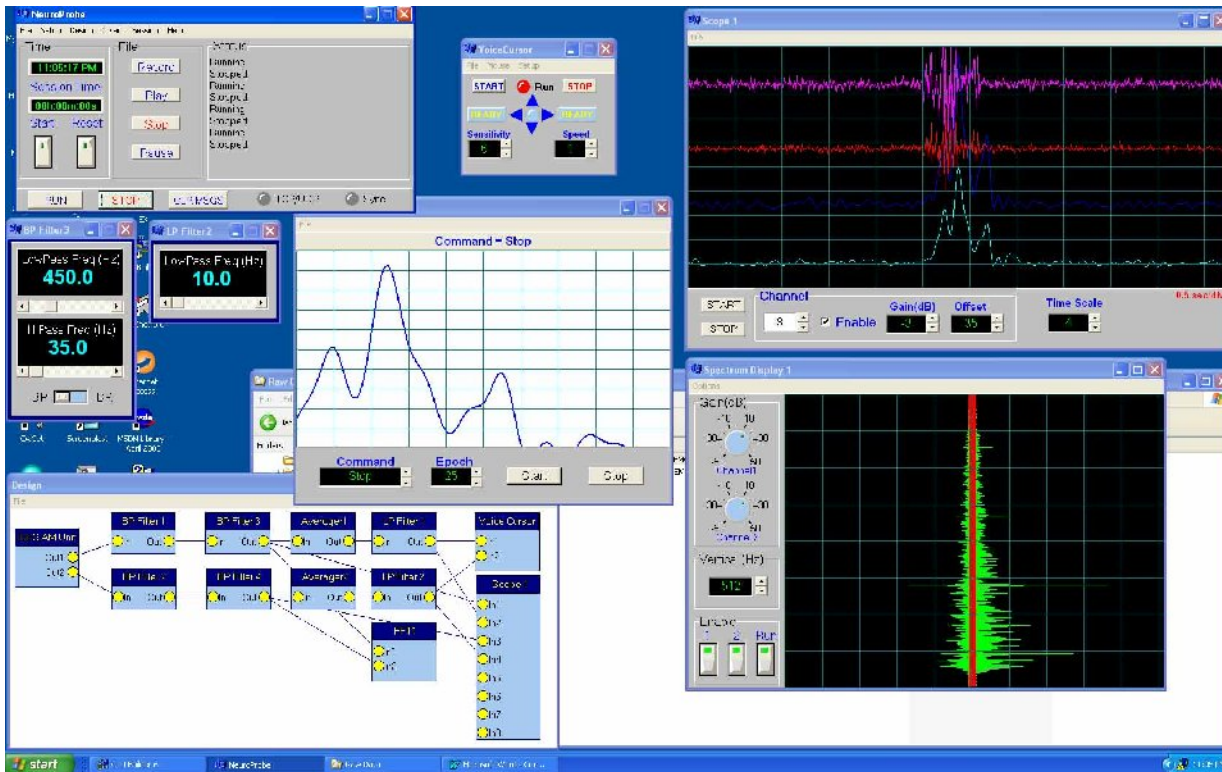


Figure 5, SoftwareTest Configuration

During initial testing the cursor and mouse clicking was disabled. However, the arrows on the Voice Cursor screen remained functional. The Vocabulary screen showed a text screen on the upper status bar depicting the selected command after the voice command input. It is also used to show the trial input number during waveform template generation.

Results; Voice Testing

There are three ways of speaking to capture EMG voice signals; speaking out loud, whispering, and speaking without using the mouth and jaw muscles (sub vocal speech). Whispering was chosen as the preferred method as it generated good EMG signals without having to use the vocal chords.

Early voice testing showed that some of the commands that were chosen have very similar templates. This led to changing the commands to create templates that were significantly different from each other to ease the recognition task. Two commands (Left and Up) that displayed similar templates are shown in the figures 6 and 7. Notice that the waveforms are normalized to about 90% of the screen height. A solution to this problem was to find an optimum location for the electrode sensors. The current location allows the chin movement (up and down) to be the dominant factor in the waveform shape.

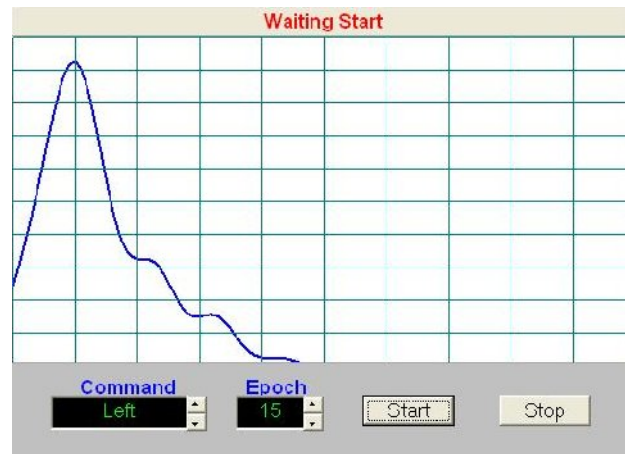


Figure 6, "Left" Command Template

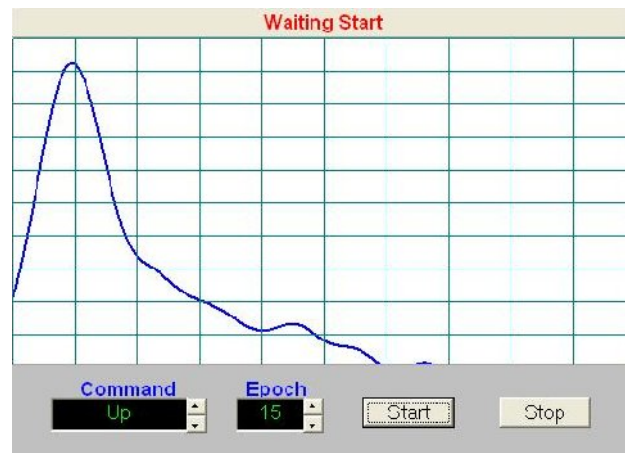


Figure 7, "Up" Command Template

Another major issue that showed up early in the testing was the inconsistency of the command input. The user has to learn to speak the command in the same manner, i. e., quickness of the command, intensity of the command, starting with the mouth open versus mouth closed, and other similar constraints. This led to poor templates being defined and/or low recognition scores even if the templates were a good

representation of the command. A solution to this problem was to define very distinct templates, like the click command, that were easy to repeat consistently. Another obvious solution was to train the user to say the command many, many times until consistent command (waveform) input was obtained.

Testing was performed to find unique templates that would be distinguished from each other and easy to repeat. Using trial and error searching, the following words were used for the commands:

Stop command: Reverse

Up command: Creamer

Down command: Down

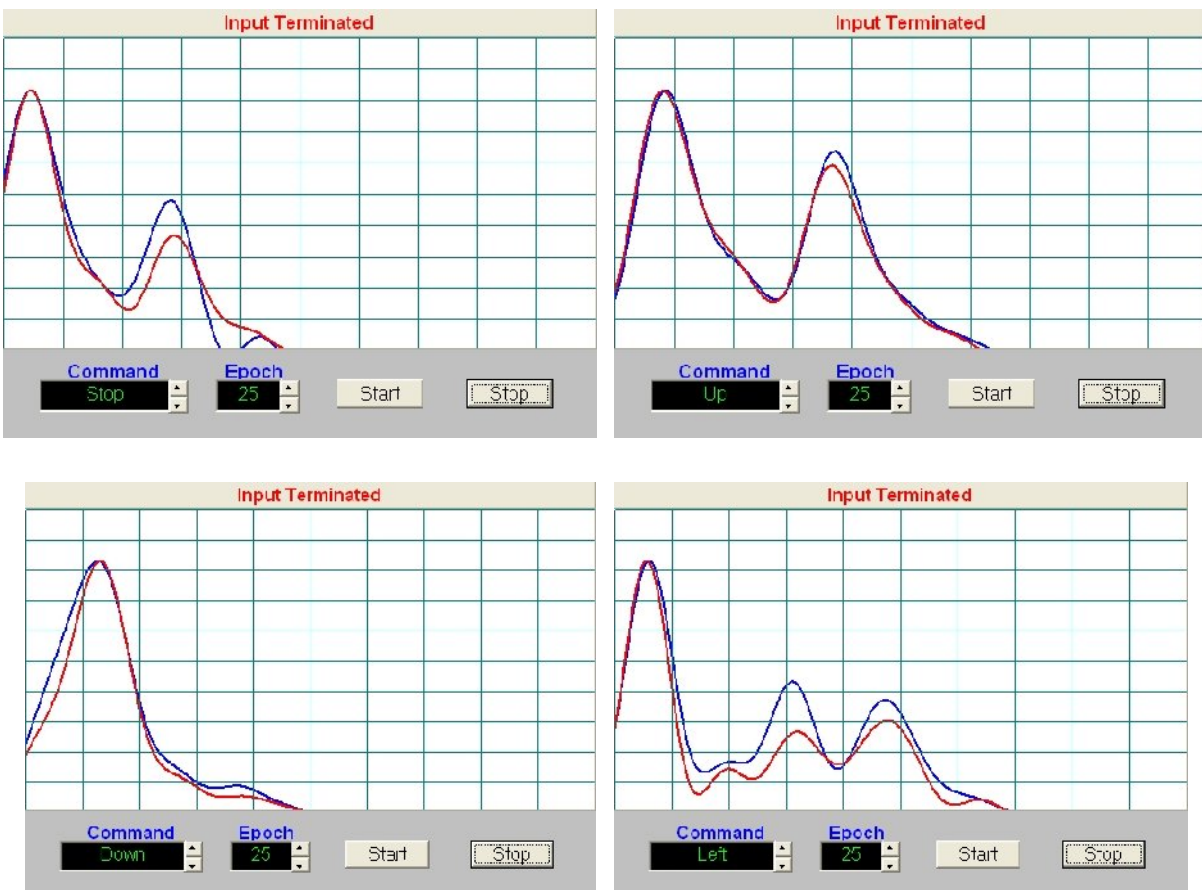
Left command: Cooler

Right command: Looky

Click command: Click

Double Click command: Click-Click

The templates for each of the commands, for an epoch of 25 tries, are shown in figure 8 below, blue = channel 1, red = channel 2:



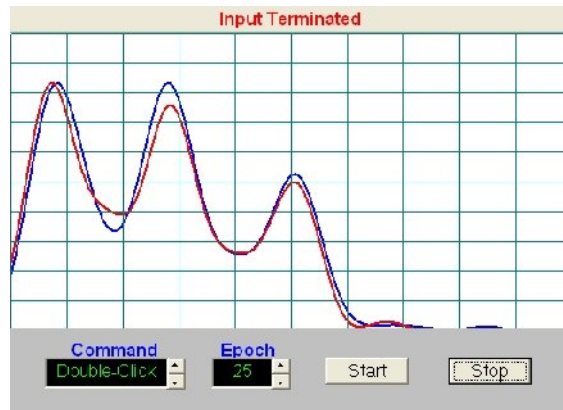
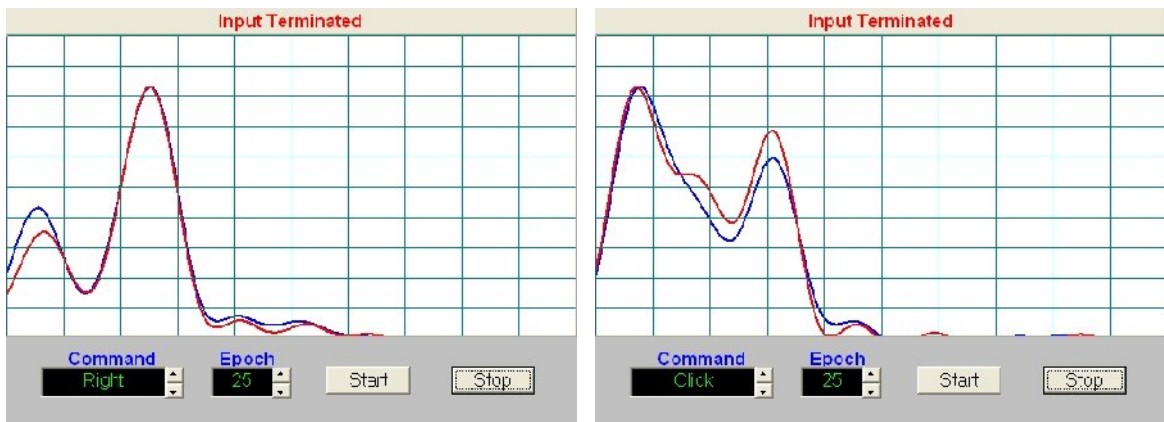


Figure 8, Command Templates

Once the templates were defined, the commands were spoken and saved in a file, each command was played back 25 times to test the recognition algorithm. NeuroProbe has a file save utility that stores the sampled data (at the specified sample rate). The file can be played repeatedly in conjunction with receiving data. The recognition algorithm had a 100% recognition rate on the saved commands. This is not unexpected since the algorithm is a "distance" based computation. Given that the noise floor has settled to the recorded value, the algorithm will produce the same results consistently.

The next voice test was to run the program with cursor and mouse clicking control. A simple keyboard application was used to input text. The cursor was run at a slow speed to allow for a learning curve in the voice control. After hours of testing it became apparent that it was too difficult to control with the DSP algorithms that used envelope detection. The derived feature sets were too variable and would not match to the templates very often. The problem was exasperated with fatigue and the non-constant/changing voice input when the frustration level increased.

Results; Facial Testing, Phase 1

Again, facial testing was performed to find unique templates that would be distinguished from each other and easy to repeat. Preliminary signal testing showed that EMG signals are "noisy" type of signals being a composite of muscle signals firing constantly and being summed at the electrodes. Figure 9 shows the EMG signal when continuously biting down. The raw and filtered EMG signals are shown for comparison. Even after filtering there are low frequency summed signals that are within the passband of the required response bandwidth. These signals were reduced through median filtering to prevent template corruption.

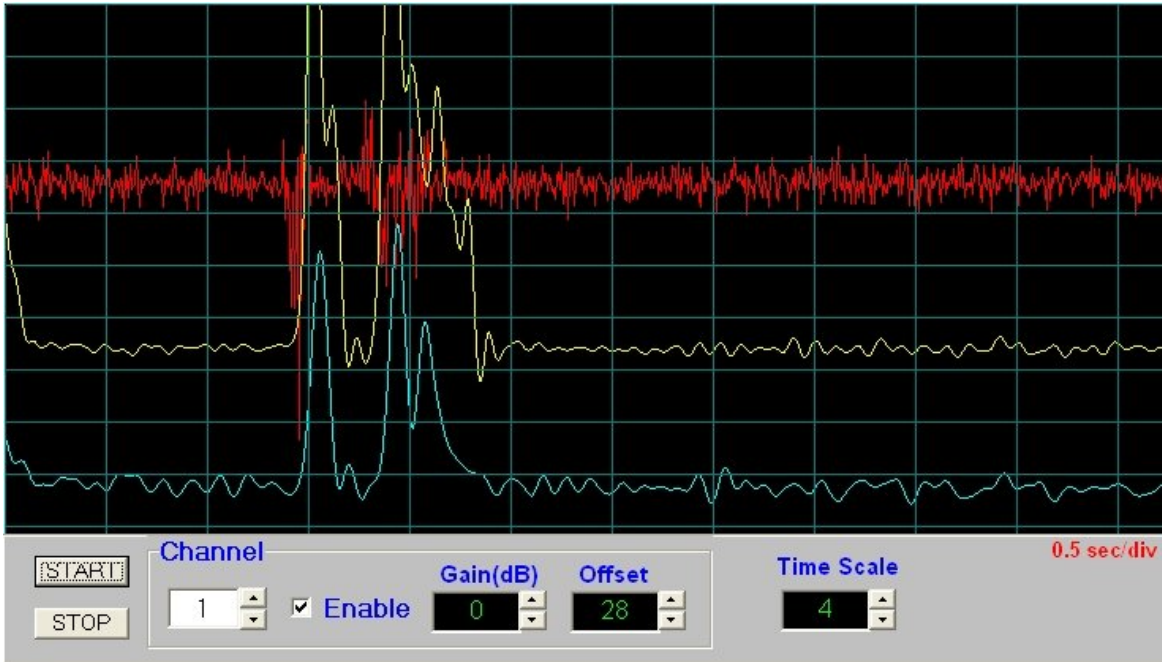


Figure 9, Facial Raw Data (top trace) and Filtered EMG Signals (Left Command)

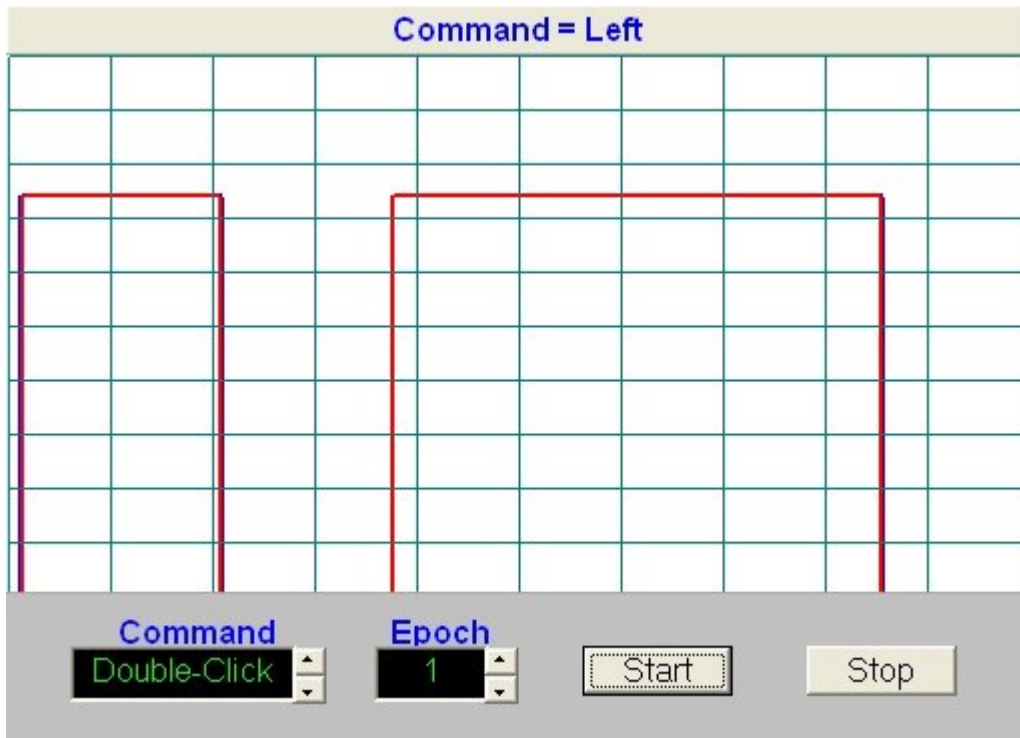


Figure 10, Left Command, Live Processed Features

It was decided to use Morse encoding to select the direction of cursor movement. The distinguishing features of the signal are pulse duration (either short or long... about a 1 to 3 ratio) and encoding of the pulse sequences. The encoding is extensible to the full Morse code set and would be useful for quick text generation. To limit the amount of command sequence generation only one and two pulse combinations were used to control the cursor. The input commands were compared to the templates over 1024 points for all cases. This required that all input had to be completed within one second. The following command codes were developed:

Stop command: dit, short pulse

Up command: dit-dit, short pulse followed by another short pulse

Down command: dah-dah, long pulse followed by another long pulse

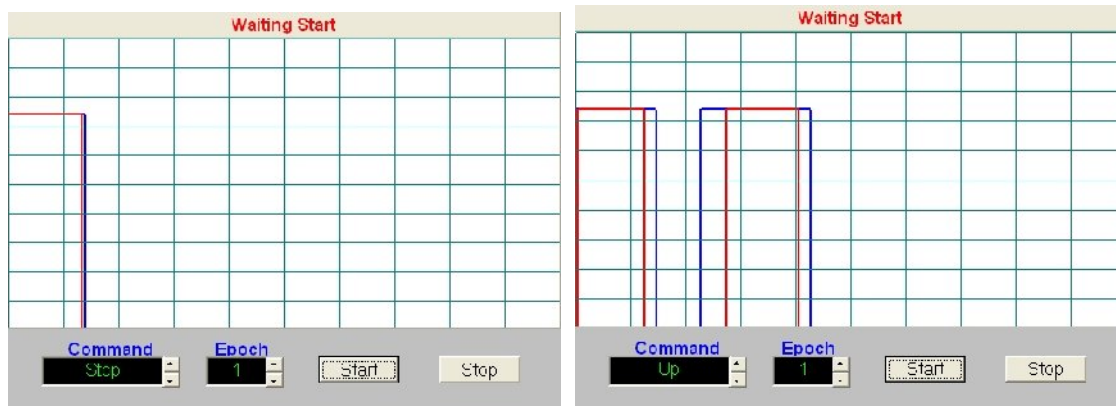
Left command: dit-dah, etc.

Right command: dah-dit, etc.

Click command: dah (four times the normal length)

Double Click command: dit-dit-dit, etc.

The templates for each of the commands, for an epoch of 25 tries, are shown in figure 11 below, blue = channel 1, red = channel 2.



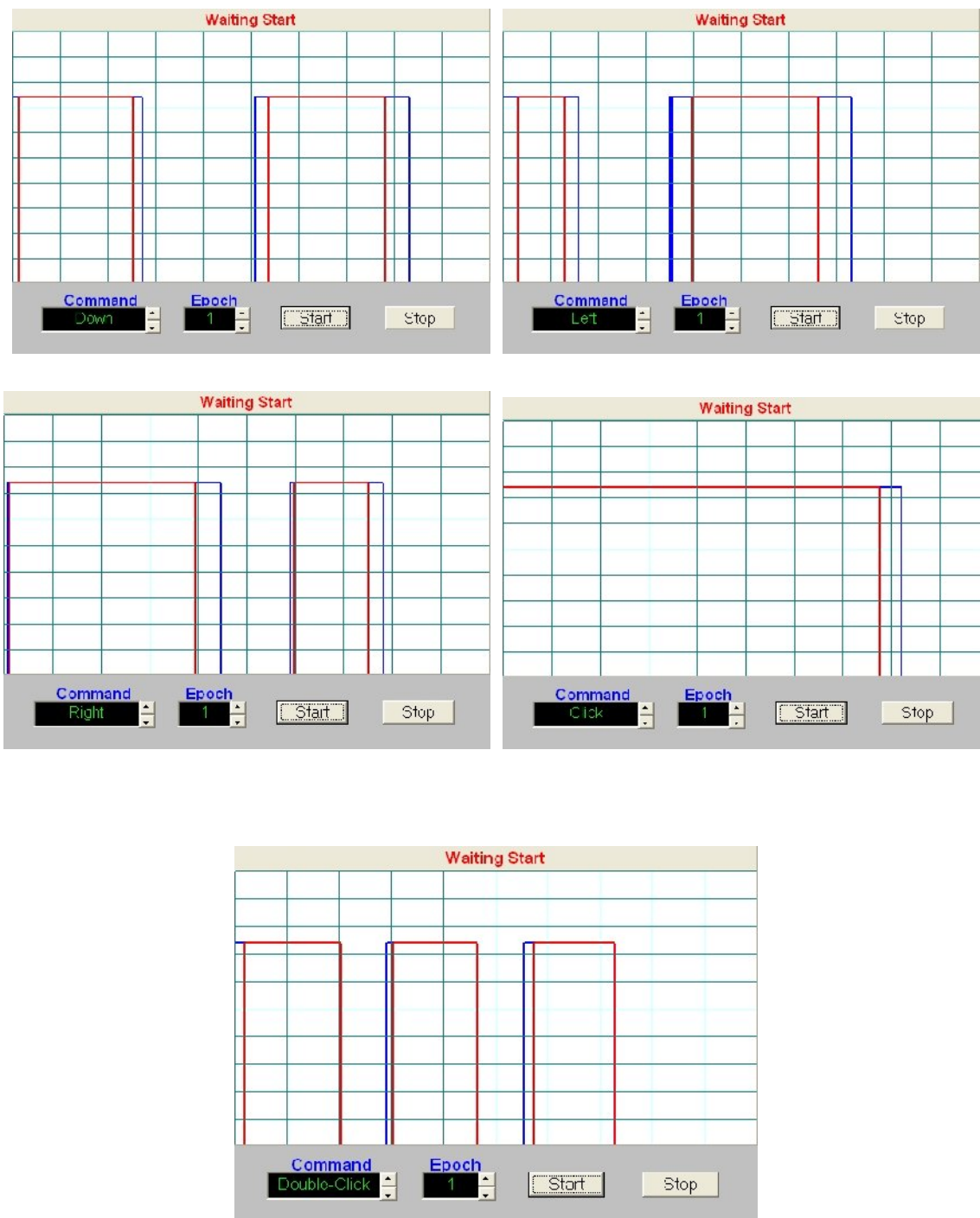


Figure 11, Facial EMG Templates

Running the cursor with the facial templates was considerably easier than with the voice input. The limit to controlling the cursor was not the recognition efficiency but rather the operator learning to create the pulses with the correct timing. There was also about a one-second delay between the input and the action response by the software. All inputs are based on a one-second data snapshot and then recognition by the software. Commands longer than a second would require minor software re-design. The control of the cursor increased with the degree of learning of the operator. This is also true for learning to key Morse code

in communication applications. Command input errors were easily terminated by quickly sending the short "stop" command. Another interesting observation was that the user was able to run for about two hours without tiring and generating excessive command errors.

Testing was performed to determine the interaction of commands. Each command was repeated 25 times by a user and logged for correctness. Some commands were unique enough that they were recognized 100%. Data that contained swallowing, coughing, or inaccurate command generation (a judgement call) were discarded. Increasing the training epoch could reduce the errors for some of the commands. Or, different commands could be used. The resultant confusion matrix is shown in table 1.

		Facial EMG Confusion Matrix					
	Stop	Up	Down	Left	Right	Click	Double Click
Stop	25	-	-	-	-	-	-
Up	-	25	-	-	-	-	-
Down	-	-	22	-	3	-	-
Left	-	-	2	21	-	2	-
Right	-	-	-	-	24	1	-
Click	-	-	-	1	1	23	-
Double Click	-	-	-	-	-	-	25

Table 1, Facial EMG Confusion Matrix

Other operational factors encountered were the starting position of the jaw, the amount of pressure of the jaw to generate a useable EMG signal, the final location of the electrodes, limiting of the input waveform for template matching, and learning Morse code (easy for some users).

VI. Conclusion

Cursor control using EMG signals was achievable and was, with practice, quite useable. The hardware, a dual channel sound card AM unit, and the NeuroProbe software were adequate for EMG signal extraction.

An attempt was made to use vocal EMG signals from the neck and under the chin. Testing showed that EMG signals were equivalent with the electrodes were placed under the chin as compared to the neck and chin. There was also a reduction in the level of the ECG signal. However, feature extraction proved to be inconsistent in real time and resulted in a low success rate for command recognition. Also many words produced feature sets that were very similar. During the search for recognizable words, it was discovered that there could be classes of words or syllables that have unique signatures. During feature extraction, these syllables produced consistent features that had stable time and energy profiles. Future work could investigate developing an alternate vocabulary/language that could be used for reliable communication. The Unique words for the commands did improve the recognition of the commands.

Facial EMG signals proved to be quite useful for establishing cursor control. Two channels were summed in amplitude and helped in reducing the noisy nature of the EMG signal. Using jaw EMG muscles was easier and less tiring than sub vocal speech. Like learning Morse code, it takes practice to space the pulses correctly in duration and in time sequence.

Future work would include developing feature DSP algorithms that work in real time for the sub vocal EMG configuration. This would open the door to many applications requiring the richness of speech input. For the facial EMG configuration the development of features for the full alphabet and numbers (including the full Morse code character set) would be desirable. This would make text input much faster than moving the mouse over a keyboard and then click with the cursor commands. A special command code could switch between cursor and text modes.

The current software operates at a sample rate of 1024 Hz. Future investigations could include operation lower sample rates (256 Hz) that would be compatible with a greater variety of EEG units. This would also reduce the computational load on the computer.

The sensors used on this project were AgCl electrodes that required electrode gel to insure a good electrical connection. In the future, capacitance type electrodes would remove the gel from the setup configuration making the device more useful and allow operation in more harsh environments.

VII. References

1. "Sub Auditory Speech Recognition" by Kim Binsted and Charles Jorgensen.
2. "Sub Auditory Speech Recognition Base on EMG/EPG Signals" by Chuck Jorgensen, Diana D. Lee, and Shane Agabon.
3. "Surface Electromyograph and Muscle fatigue" by Eric Beltt, May, 2002.
4. "Small Vocabulary Recognition Using Surface electromyography in Acoustically Harsh Environment" by Bradley J. Betts and Charles Jorgensen, NASA/TM-2005-21471.
5. "Web Browser Control Using EMG Based Sub Vocal Speech Recognition", By Chuck Jorgensen and Kim Binsted.
6. "Decomposition of Surface EMG signals", by Carlo J De Luca, Alexander Adam, Robert Wotiz, L. Donald Gilmore, and Hamid Nawab. January 2006.

Notes:

Figures 1 and 2 are Copyright © 2006, Les Laboratories Sevier