Gesture Control By Wrist Surface Electromyography

Abhishek Nagar and Xu Zhu
Samsung Research America - Dallas
1301 E. Lookout Drive
Richardson, Texas 75082
Email: {a.nagar, xu.zhu}@samsung.com

***Abstract*—Surface electromyography (SEMG) systems are able to effectively sense muscle activity, irrespective of any apparent body motion, in a highly convenient and non-intrusive manner. These advantages make SEMG based systems highly attractive for use as a human computer interface. Despite such advantages, there are still a significant amount of challenges that should be resolved before such systems can be made viable. In this paper we focus on a wrist based SEMG system that is required to detect as well as recognize the gesture being made by the user. A major challenge in the detection of a gesture in an SEMG signal is the noise due to displacement of electrodes on the skin which does not belong to any of the well studied noise types. We use a bilateral filtering based approach to estimate such noise and then effectively detect the gesture signal. Next, we identify the gesture based on information contained in different frequency bands of the signal. Based on our experiments, we show that our system achieves an accuracy of 88.3% in identifying the correct gesture among *rock*, *paper*, and *scissors* gestures.**

I. Introduction

With the proliferation of computers in our daily lives, it is now ever more important to have an effective means to control or operate a computer. Two of the important qualities of an effective computer input device that have been identified in literature are: i) Natural so that it does not consume a significant portion of user's attention while at the same time effectively conveying information to the computer, and ii) Unnoticeable so that it does not create disturbance in user's surrounding [1]. These qualities are very much present in surface electromyography (SEMG) thereby making it a promising contender for designing a computer input device.

Surface electromyography essentially captures the activity of muscles using a set of electrodes placed on the user's skin and is thus able to effectively read any contraction or relaxation of the associated muscle irrespective of the fact that whether that contraction or relaxation actually resulted in any body movement. This ability of an SEMG based system to read the intentions of a user allow it to be used as a highly effective computer input device. See Section II for a detailed discussion on the use of SEMG system as a computer input device. A typical SEMG based computer input device used in research is in the form of an armband that is able to detect various gestures made by the arm or the hand based on the captured SEMG signals [2]. With the maturing of this technology, there is increased interest in the commercial sector as well to commoditize this technology. A recent startup company called Thalmic Labs Inc. is already producing SEMG based gesture recognition systems in the form of wearable armbands [3]. See Figure 1. One of the limitations of the devices such as Myo is that they need to be worn on the arm very close to the elbow in order to cover a larger muscle area. This could be very inconvenient and unnatural. In this paper, we use a wrist surface electromyography (SEMG) device which can be worn as a typical wrist band worn during athletic activities and provides the same functionality as Myo. Figure 2 shows our current experimental SEMG wrist band.

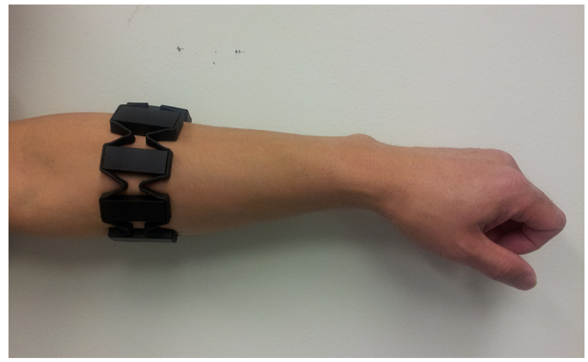


Fig. 1. Myo: An SEMG based gesture recognition device from Thalmic Labs Inc. [3].

Fig. 2. Our experimental setup: Wrist SEMG band with two signal channels configured using four electrodes.

In this paper, we address two main aspects of the SEMG based gesture recognition system: i) gesture detection i.e. segmentation of SEMG signal into regions where the gestures were being made and regions where no gesture was being made, and ii) gesture recognition i.e. identifying which gesture is being made. Owing to the highly uncontrolled noise present in the SEMG signal due to the slippage of the SEMG electrodes on the skin, gesture detection is not straightforward. We use bilateral filtering based approach and show that it is highly effective in robustly segmenting the SEMG signal. Next, we extract certain salient features from the SEMG signal based on different frequency bands that allow accurate recognition of the type of gesture being made.

II. Surface Electromyography (SEMG) based Gesture Recognition

The skeletal muscles contract or relax based on the electrical impulses or action potential provided by the motor neurons attached to the muscle. Measurement and analysis of such action potentials applied on muscle fibers is referred to as electromyography[4], [5]. In order to obtain a clean electrical signal from a muscle, needle shaped electrodes are typically inserted into the specific muscle. Need for such assessment frequently arises in the field of medicine, e.g. to assess muscle tone and strength during a kinesiologic analysis, diagnosis of neuro-muscular disorders, and prosthetic limb design. However, one advantage of electromyography is that the EMG signals can also be captured from the skin using surface electrodes. The signal thus captured is significantly more noisy than the EMG signal captured from the needle electrodes due to i) impedance from the layer of fat and skin, ii) interference of signals from multiple muscles, and iii) poor contact of the surface electrodes to the skin. Despite these drawbacks, surface electromyography systems that capture the EMG signal from the skin surface are very popular as these significantly reduce the discomfort of the subject and avoid any chance of infection caused due to the inserted electrode. With the greater convenience afforded by the SEMG systems, these are now commonly used for gesture recognition. Due to their property of being able to detect subtle gestures, these systems are referred to as "intimate interfaces" [1], [6].

Saponas et al. [2], [7], [8] did an extensive work on developing SEMG based gesture recognition systems. In [2] a thorough introduction of SEMG based systems is provided and an in-house SEMG based gesture recognition system is developed in the form of an armband. Extensive experiments were provided based on this device to show the feasibility of four different types of gestures: i) Position: placing index or middle finger at two different positions on a surface, ii) Pressure: applying low or high pressure on the index or middle fingers, iii) Tap: tapping with either of the five fingers, and iv) Lift: lifting either of the five fingers in air. Here the subjects were given visual instructions about when to make the gesture and when to not thereby avoiding any automatic signal segmentation. In [7] Saponas et al., extended their work in [2] by incorporating more natural gestures including ones that were made while the user is either holding a travel mug or holding a bag. Here, in order to detect when the user made the gesture, two SEMG electrodes were placed on the "other" arm not making the gesture and the user was asked the strain this "other" arm when making the gesture. Note that straining the arm can be easily captured as sudden jump in the SEMG signal. In [8], Saponas et al., extended their work in [2], [7] by avoiding use of adhesive or gel on the electrodes while collecting the data and avoiding re-training the system every time the SEMG armband is worn. They show that the designed system is 86% accurate in identifying one of the three fingers used in making the pinch gesture. Here they used data from six SEMG signal channels using silver chloride electrodes for classification.

III. Hardware Design

Ideally, if the SEMG system is required to sense EMG signal from multiple muscles, multiple electrodes are placed on different muscles. However, in our case since the muscles are very narrow near the wrist where they are essentially being connected to the bone area and thus the signal may be weak, we use a large sensor to cover most of muscles instead of trying to probe individual muscles. Although the cross coupling of signals from multiple different muscle is expected to be strong here, with signal processing procedure described below, certain gesture can still be successfully recognized. Due to the limited space to place the electrodes, the wrist band uses single end signal input setup instead of commonly used differential configuration. With this setup, we place two signal channels on the wrist band conveniently without placing the electrodes very close to each other. As a common practice in EMG setup, reference electrode for two single end channels is place right on the wrist flat bone area and the RLD (right-leg drive) is placed on the side of wrist with least amount muscle movement to reduce the 60Hz interference. We use a wireless analog front end (AFE) device that communicates with the computer via a bluetooth interface. The AFE chip used in our system is TI ADS1299. The key parameter of system are listed in the table I

Attribute	Value
Number of bit for ADC	24bit
Sampling frequency	500Hz per channel
Input-Referred Noise	1.0 V
Common Mode Reject Ratio	110 dB
Programmable Gain	1, 2, 4, 6, 8, 12, or 24
Input signal range	Max +/- 100mV
Max input channel	8 (2 used in this work)

TABLE I. Key performance for sEMG hardware AFE.

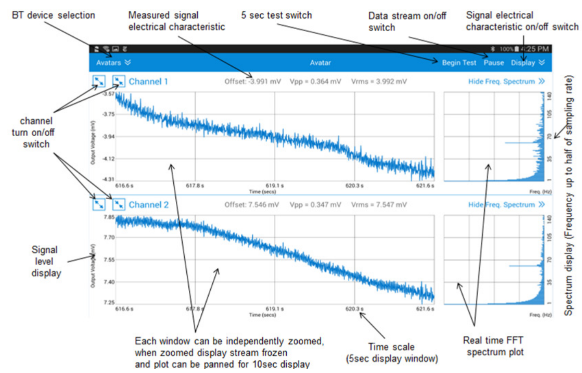


Fig. 3. Data acquisition interface built on Android 4.4 platform.

The data acquisition software is developed on Android 4.4 is illustrated in the figure 3. The software depicts in real-time the two SEMG channels being captured from the device. This software is very useful in checking if the electrodes are properly in contact with the skin and are connected to the capture device and also to indicate to the user when to make a specific gesture. This software pairs with the SEMG hardware via bluetooth.

IV. Signal Analysis

There are two main sources of noises that affect an SEMG signal: the interference of the external electro-magnetic field and electrode movement on skin that is further accentuated due to the high skin impedance [9].

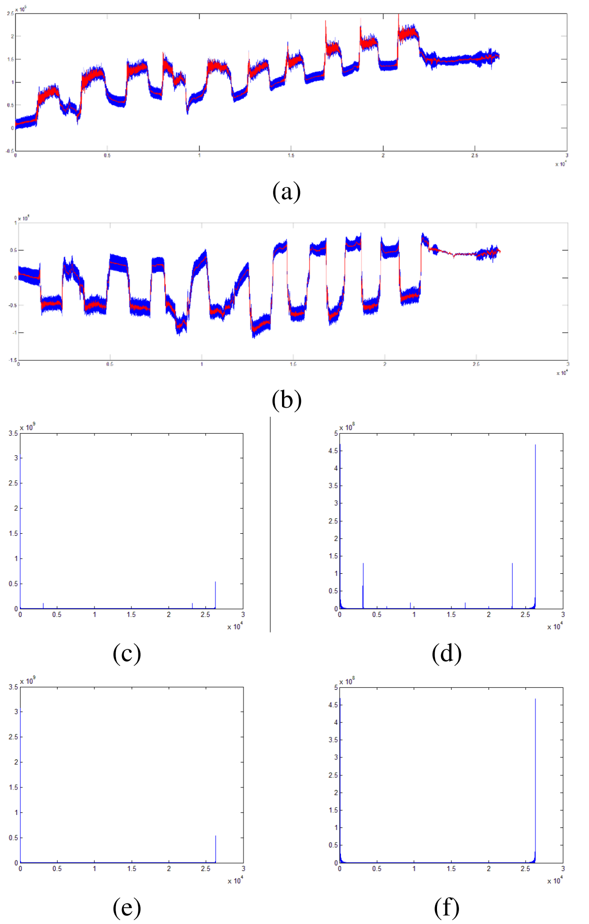


Fig. 4. IIR notch filtering for $60, \quad 120, \quad 180, \text{and} \quad 240 Hz$. (a) shows the original (blue) and filtered (red) signals from channel-1, (b) shows the original (blue) and filtered (red) signals from channel-2, (c) shows the magnitude of Fourier transform of channel-1, (d) shows the magnitude of Fourier transform of channel-2, (e) shows the magnitude of Fourier transform of channel-1 after filtering, and (f) shows the magnitude of Fourier transform of channel-2 after filtering.

A. *Interference noise removal*

There is significant amount of literature addressing the noise due to external interference. In order to eliminate the interference noise we use the second order IIR notch filter tuned to $60Hz$ and its harmonics. Since our signal was captured at a frequency of $500Hz$, we use a set of four notch filters tuned to $60, \quad 120, \quad 180, \text{and} \quad 240Hz$. As per our experiments, these filters are highly effective in eliminating the undesired interference noise. Figure 4 shows the results corresponding to the use of notch filters.

B. *Offset noise removal*

Since skin is a soft tissue, it is not possible to ensure that the electrodes remain at exactly the same location with respect to the muscle during their use. This motion of the electrodes causes significant change in the electrical potential measured by the electrodes. This noise manifests itself in the form of an offset to the captured SEMG signal. Since the actual SEMG signal is expected to zero mean, a simple approach to eliminate the offset is to estimate the noise as the moving average of the signal. However, one of the major difficulty is that the noise signal can be arbitrarily smooth or fluctuating making it difficult to model it and eliminate from the main signal. In order to address this issue, we utilize bilateral filtering based approach to estimate the trend of the signal. See Section V for a description of the bilateral filtering approach.

In our case, however, since the signal is one dimensional the above analysis would be relevant if we assume the height of the image to be 1 pixel. In our context since we are trying to estimate the trend of the signal that is dominated by the offset noise, we can consider the pure SEMG signal to represent the white noise in the image that need to be removed and the large abrupt changes in the offset noise would represent the strong edges in the image to be preserved. In order to de-trend the signal, we first estimate the trend line as the bilateral-filtered signal and the final processed signal is obtained by subtracting the bilateral-filtered signal from the original signal, i.e.

$$X_1 = X - \mathcal{F}_b(X) \tag{1}$$

where X is the original signal, X_1 is the processed signal, and $\mathcal{F}_b$ represents bilateral filter operator.

C. *Signal segmentation*

Given the SEMG signal with offset noise removed as depicted in Eq. 1 by X_1, we compute the moving average of $|X_1|$ as a measure of local variability of the signal i.e.

$$Y = \mathcal{M}(|X_1|) \tag{2}$$

where $\mathcal{M}$ is the moving average operator. In our experiments, $\sigma_s = 10$, and $\sigma_r = 20000$ and for moving average, the width of filter is 100 sample points.

Since we capture data from two different SEMG channels, to have a single indicator vector for the two channels, we sum the moving averages from the two signals to obtain a single signal. We then subtract the minimum value of the signal from itself and threshold it using its mean value to obtain a binary signal indicating where the gesture was made. Figure 5 shows the different stages of gesture segmentation.

V. BILATERAL FILTERING

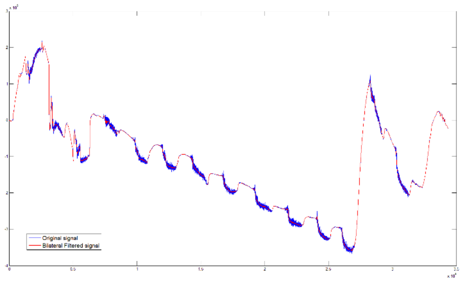
Bilateral filter is an edge preserving smoothing filter for images. In bilateral filtering, the filtered value of a pixel is computed as weighted average of the neighboring pixels where the weight for the neighbors are also determined by the difference in their intensity values from the center pixel. More specifically,

$$I^{filtered}(x) = \frac{1}{W_p} \sum_{x_i \in \Omega} I(x_i) f_r(||I(x_i) - I(x)||) g_s(||x_i - x||) \tag{3}$$

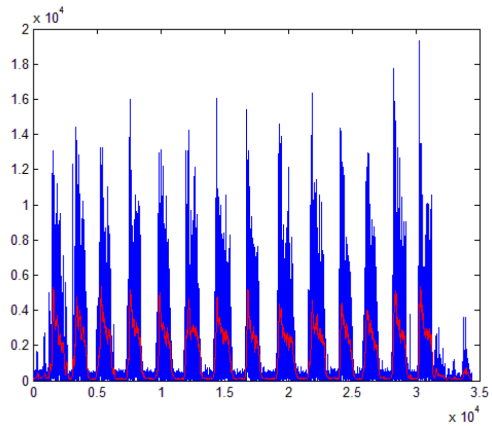
where

$$W_p = \sum_{x_i \in \Omega} f_r(||I(x_i) - I(x)||) g_s(||x_i - x||). \tag{4}$$

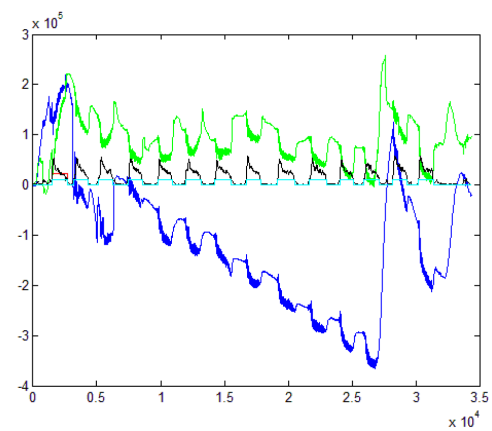
Here, $I^{filtered}$ is the filtered image, I is the input image, x are the coordinates of the current pixel, Ω is the window



(a)



(b)



(c)

Fig. 5. Gesture segmentation procedure. (a) shows original signal (blue) and bilateral filtered signal (red), (b) shows the fully rectified signal obtained after subtracting the bilateral filtered signal from the original signal, also described as $|X_1|$ in Equation 1 in blue and its moving average which is described as Y in Equation 2 in red, and (c) shows the raw data from the two channels: channel-1 in green and channel-2 in blue, the combined variability indicator for the two channels in black and the segmentation indicator in cyan. The red portion in the segmentation indicator indicates the portion of the segment that was discarded based on the selection rules defined in VII-B.

Original image Filtered image

Fig. 6. Bilateral filtering.

centered on x, f_r is the range kernel (e.g. Gaussian) for smoothing differences in intensities, g_s is the spatial kernel (e.g. Gaussian) for smoothing differences in coordinates,

$$f_r(a) = \frac{1}{\sqrt{2\pi}\sigma_r} e^{\frac{-a^2}{2\sigma_r^2}}, \text{ and} \tag{5}$$

$$g_s(a) = \frac{1}{\sqrt{2\pi}\sigma_s} e^{\frac{-a^2}{2\sigma_r^2}}. \tag{6}$$

Here σ_r determines the intensity range of the filter, increasing σ_r leads to reduced sensitivity to any variation in intensity thereby making the filter similar to a Gaussian-blur filter. The second parameter σ_s determines the spatial spread of the filter, increasing σ_s will lead to increased smoothing over a large support. Figure 6 shows an example image and its filtered version using bilateral filter.

VI. Gesture Classification

Once the signal segments have been detected, the signal portion can then be classified as one of the three gestures. We essentially use frequency based features for this task.

A. Feature Extraction

Given the different segments of the signal, we first obtain a spectrogram [10] of the signal. Note that a spectrogram is essentially a Short-time Fourier Transform of the consecutive overlapping partitions of the input signal. The two parameters for a spectrogram are the size of partition and size of the overlap. In our experiments, we use partitions with 513 sample points and the consecutive segments have an overlap of 413 samples. In order to obtain the spectrogram, each partition is padded with zeros to upto a total length of 1024. Short-time Fourier transform is performed on this signal and due to the symmetry in the Fourier domain signal, first 513 samples are returned as the spectrogram component of that partition. In our experiments, we only consider the absolute value of the spectrogram. We extract four different types of features from each partition for classification:

1) **Low frequency features**: This feature captures the energy in the low frequency component of the spectrogram. Specifically it contains the sum of first 30 elements of the Fourier transformed signal for both the channels.
2) **Channel difference feature**: This feature captures the energy in the difference between the two channels

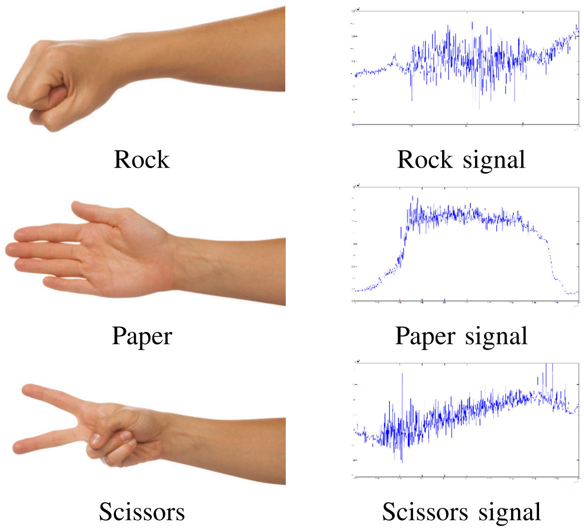


Fig. 7. Three gestures used in our experiments.

and it is thus computed as the sum of the signed difference between the two Fourier transformed signals corresponding to the two channels.

3) **High frequency features**: This features captures the energy in the high frequency component of the spectrogram. Specifically it contains the sum of last 256 elements of the Fourier transformed signal for both the channels.

4) **Intermediate frequency aggregates**: This feature is a multi-valued feature which captures the characteristics of the intermediate frequency bands with 11 frequency bands each of size 20 from 30^{th} element to 240^{th} element. Here we capture 3 different types of features: in the first case, we capture the above mentioned 11 features for the Fourier transform for the first channel, the second case is for the second channel and the third case is for the signed difference between the two channels. In all, this feature provides a vector of length 33 for each partition.

B. Classifier Design

Thus, in all there are 36 features for each partition of the spectrogram. In order to extract further discriminative features from these 36 features, we use pairwise Linear Discriminant Analysis (LDA) [11] features to reduce its dimensionality to 3 with each dimension discriminating among one of the three pairs of classes. Given a training dataset, we first fit a multivariate Gaussian distribution on features from each of the three types of gestures and during testing, we classify a feature to the class corresponding to which its probability density is the highest.

VII. Experiments

For our experiments, we used signal from two different subjects and from both left and right wrists. We used rock, paper and scissors as the representative gestures. Figure 7 shows the three gestures and three sample SEMG signals corresponding to the three gestures.

A. Data Acquisition Protocol

During the data capture, the subject is asked to sit comfortably on a chair with the electrode wrist band worn. In order to have good electrical contact of the dry silver chloride coated electrodes with the skin, we apply some electrolyte gel on the skin. Then the software is connected via bluetooth to the device and the signal is visually checked to see if there is sufficient variability in the signal when a gesture is being made. Once the wrist band is ready to be tested, the subject is asked to focus on the tablet running the data acquisition software and when the screen shows a textbox reading "Make Gesture", the subject is asked to make the gesture asked. The software is configured to display the "Make Gesture" signal for 2 seconds with 3 seconds gap in-between gestures. Once the wrist band is ready to be tested, the subject is asked to make a said gesture for approximately 10 times. We collect approximately 10 instances of gesture for each type.

B. Data Post-processing

Many a times the user inadvertently stresses the muscle leading to a spike train in the captured signal. However, since the time-span of these short signals can be easily distinguished from the required time span of the gesture, these can be easily filtered. For this, we first fill any small discontinuities in the binary indicator vector as mentioned in Section IV-C and then we verify the time spans of the gestures and eliminate gesture spans small than or larger than certain set thresholds. Further, based on the instructions provided to the user, we ensure that the left and right gaps around a gesture signal are not smaller than half the length of the gesture and not larger than twice the length of the gesture in order to avoid incorrect gesture detection.

C. Results

In order to validate our technique, we visually inspected our segmentation results for a large number of candidates and the portions of the signal segmented as gesture region was most of the time correct. Figure 8 provides a number of sample instance for gesture detection.

Next we extracted the spectrogram based features and classified each frame of the spectrogram using the technique described in Section VI. In order to classify a contiguous gesture signal, we took the majority vote among the classes assigned to each of the frames inside the contiguous signal. Based on this analysis, we observed that the system could obtain a recognition accuracy of 88.3% based on 10 fold cross validation.

VIII. Conclusion

In this paper, we have studied the problem of gesture detection from a two channel SEMG source and gesture recognition on the same. For gesture detection, we have provided qualitative results whereas for gesture recognition, we provide quantitative results and show that a wrist based gesture recognition system is viable based on current technologies. As future work, we plan to further extend our experiments with more data and quantitative evaluation of gesture detection as well as recognition. We also plan to study other aspects of the wrist SEMG based gesture detection such as the effect

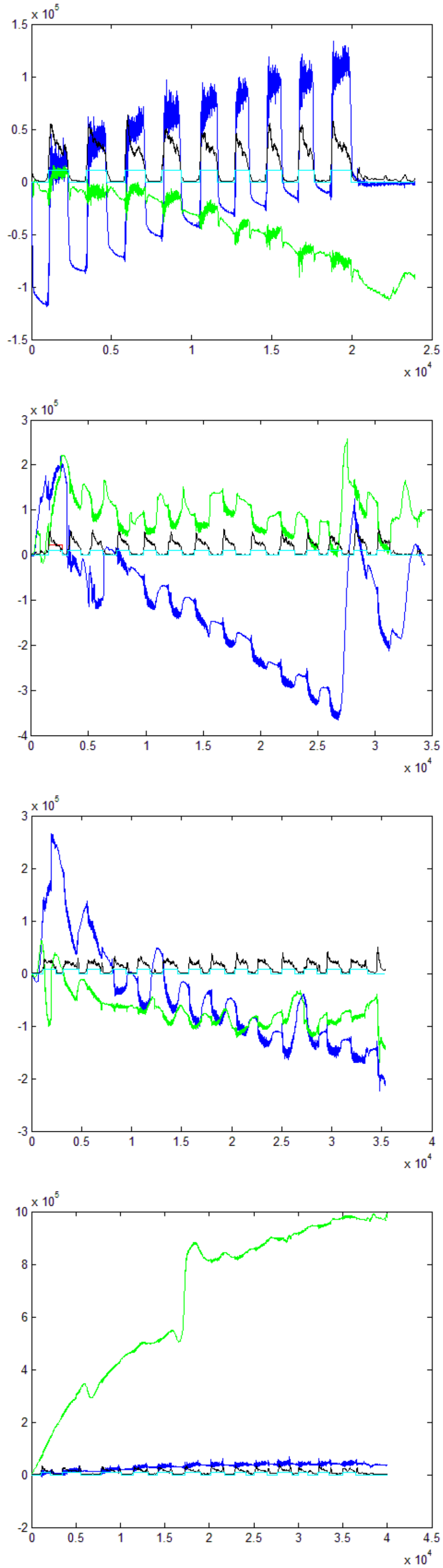


Fig. 8. Sample gesture detection results.

displacement of electrodes, effect of demographics such as age, race, gender etc.

REFERENCES

[1] E. Costanza, S. A. Inverso, and R. Allen, "Toward subtle intimate interfaces for mobile devices using an emg controller." in *ACM CHI Conference*, 2005, pp. 481–489.

[2] T. S. Saponas, T. D., D. Morris, and R. Balakrishnan, "Demonstrating the feasibility of using forearm electromyography for muscle-computer interfaces." in *Proceedings of the 2008 SIGCHI Conference on Human Factors in Computing Systems*, Florence, Italy, April 2008.

[3] "Thalmic labs inc." https://www.thalmic.com/en/myo/.

[4] J. V. Basmajian and C. DeLuca, *Muscles Alive*, 5th ed. Baltimore, MD: Williams and Wilkins, 1985.

[5] E. Criswell, *Crams Introduction to Surface Electromyography*. Jones and Bartlett Learning, 2010.

[6] E. Costanza, S. A. Inverso, R. Allen, and P. Maes, "Intimate interfaces in action: Assessing the usability and subtlety of emg-based motionless gestures." in *ACM CHI Conference*, 2007, pp. 819–828.

[7] T. S. Saponas, D. S. Tan, D. Morris, R. Balakrishnan, J. Turner, and J. A. Landay, "Enabling always-available input with muscle-computer interfaces," in *Proceedings of the 22nd annual ACM symposium on User interface software and technology*, BC, Canada, October 2009.

[8] T. S. Saponas, T. D., D. Morris, J. Turner, and J. A. Landay, "Making muscle-computer interfaces more practical." in *Proceedings of the 2010 SIGCHI Conference on Human Factors in Computing Systems*, Atlanta, GA, April 2010.

[9] C. J. DeLuca, L. D. Gilmore, M. Kuznetsov, and S. H. Roy, "Filtering the surface emg signal: Movement artifact and baseline noise contamination," *Journal of Biomechanics*, vol. 43, p. 15731579, 2010.

[10] B. Boashash, Ed., *Time Frequency Signal Analysis and Processing: A Comprehensive Reference*. Oxford, U.K.: Elsevier, 2003.

[11] O. Duda, P. Hart, and D. Stork, *Pattern classification*. John Wiley & Sons, 2000.