Final Project Report 18-551, Spring 2012 Expectation-Maximization Guidance (EMG)

Group 6

Ben Cowley - bcowley@andrew.cmu.edu

Benjamin Shimomura - bshim@cmu.edu

INTRODUCTION

Prosthetics for hand and lower arm amputees are limited in usefulness. Currently, a prosthetic hand would not allow a human to type, to use a touch screen, or to otherwise interact with a computer, especially handheld phones, which require the dexterity to hold it while touching a relatively small screen. However, amputees do retain the nerves and muscles required for human-computer interfaces, so they are still able to send action potentials down toward their amputated arm, resulting in signals recorded as EMG.

A useful human-computer interface should not be difficult to understand; similarly, an intuitive human-computer interface can lead to potential profit. For example, the rise of graphical user interfaces (as opposed to command line interfaces) dramatically increased the accessibility and perceived usefulness of personal computers. EMG itself requires an intuitive interface, since many users have built up decades of “muscle memory” and hand-eye coordination. If EMG-based interfaces are to be of any consequence, they must be user-defined.

Electromyographic (EMG) signals are currently used non-medically for gesture recognition [1, 2], general signal classification [3], and biometric recognition [4]. EMG data is also used in a wide range of medical fields of study, including neuromuscular diseases, kinesiology, and motor control disorders [5]. EMG signals are typically collected using either surface or needle electrodes [6], but can also be recorded using certain conductive textiles [7].

BIOMETRICS

Prior Work

Biometrics have been primarily considered in at least two prior groups: “Identity Recognition” (S11, G2), and “Facebook Tagging” (F09, G1). Both of these groups focused on face recognition, and achieved results of ~85% recognition among 20 individuals, albeit with a very restricted training and testing data set (“Facebook”) and ~43% recognition among 9 individuals, with fewer restrictions on the kind of acceptable data.

However, neither of these dealt with using EMG as a biometric. In “Super Hand” (S11, G3), Lin and Kator discuss biometric application of EMG data, but candidly, because electrodes were placed in different locations, it is hard to ascertain whether good subject identification for some hand motions were a result of biometric differences or a confounding variable in the experiment.

Our work hopes to improve on these prior 18551 projects by leveraging EMG data, a dynamic source of biometric features, to distinguish subjects, rather than static face data. While Lin and Kator tested on multiple motions, we test on a single “motion”—keyboard-based “password” entry. Additionally, we segment using keystroke data rather than manually segmenting each trial. Finally, we increase the number of subjects as compared to “Super Hand” and attempt to fix electrode placement in order to ascertain feasibility of EMG data as a reliable biometric.

EMG data has not often been studied as a possible biometric. Prior work, to the best of our knowledge, is limited to work by Suresh, *et al*, who were able to achieve 98% recognition among 49 subjects, when testing and training on the same experimental session. They achieved 86% recognition when testing and training on different experimental sessions. Testing was based on a restrictively defined upward wrist motion. They conducted experiments with one differential pair of electrodes. A Gaussian mixture model, implemented using the EM algorithm, was used to better model the likelihoods of the EMG-based features, as opposed to a vector quantization model. Features were extracted by passing the signal spectrum through a non-uniform filter bank.

Our work adds to this previous work by increasing the number of electrode differential pairs to 8 (eight), significantly altering the experimental procedure to one of typing a password, decreasing the complexity of our classification algorithm, and segmenting using keystroke data.

We are grateful to the lab of Marios Savvides for their help and support, and in this report we will only discuss work to which we originally contributed.

We generated all MATLAB code. Keystroke data was obtained through keylogging software provided by the lab of Marios Savvides.

Data

30 sessions were recorded from 14 subjects. Exactly 3 sessions were recorded from 2 subjects; exactly 2 sessions were recorded from 12 subjects. Approximately 100 and 50 trials (including incorrectly typed strings) were recorded during each subject's first and second session, respectively.

Subjects typed the string "Hello, world.#18" followed by the [Enter] key using a standard USB keyboard. On each key press and key release, the internal computer time (in ms) along with the key pressed or released was recorded. The surface voltage of each subject's hands and forearms was recorded in mV at 960 Hz with a maximum voltage of 3 mV using the BioRadio. The BioRadio by Cleveland Medical Devices, Inc., including its wireless radio to serial-USB receiver, were used along with the BioCapture Lite software. 17 snap MVAP II Electrodes from MVAP, with 17 insulated snap leads, were used to measure 8 channels (1 was used to ground the BioRadio). Digital 5 Hz high-pass filters with 60 Hz notch filters were implemented on each channel with the BioCapture Lite software.

Electrode placement is shown in Figure 1. Electrodes were placed on the skin above the hypothenar eminence, the thenar eminence, the extensor carpi ulnaris and anconeus, and the flexor carpi radialus and palmaris longus. Two pairs were placed on the hands. The electrode pair recording the electrical activity of the thenar eminence was intended to record much of the EMG data from the first finger, while the pair recording the hypothenar eminence was intended to record primarily the little "pinky" finger. The other two pairs were placed approximately on the anterior and posterior of the forearm. The electrode pair recording the carpi ulnaris and anconeus was intended to record the third "ring" finger. The pair recording the flexor carpi radialus and palmaris longus was intended to record the second "middle" finger.

Electrodes were placed by inspection of which muscles are most prominent near the skin while the subject makes typing motions. From a seated position, a subject's arms are bent at the elbow in order to type at a keyboard, which narrows the number of prominently active muscles. The visible flexing of the muscles implies that electrodes placed near these muscles will record a strong EMG signal, though not necessarily a biometrically meaningful signal.

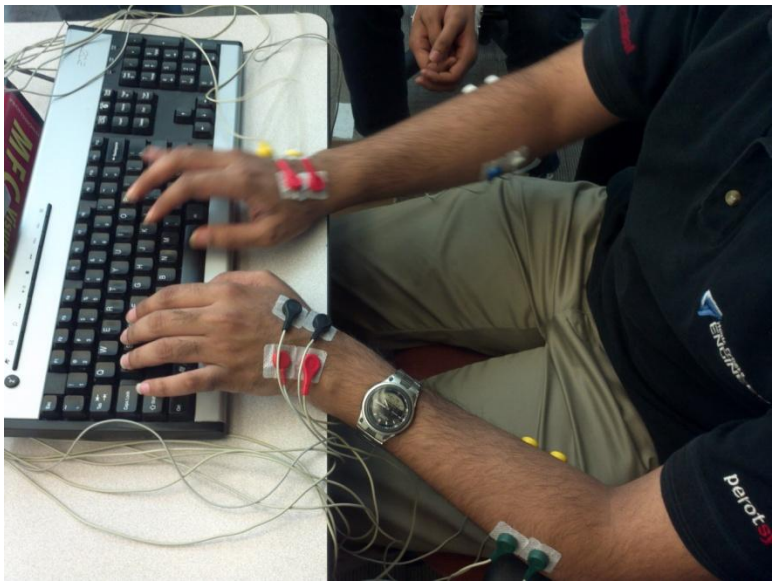

Figure 1 (a).

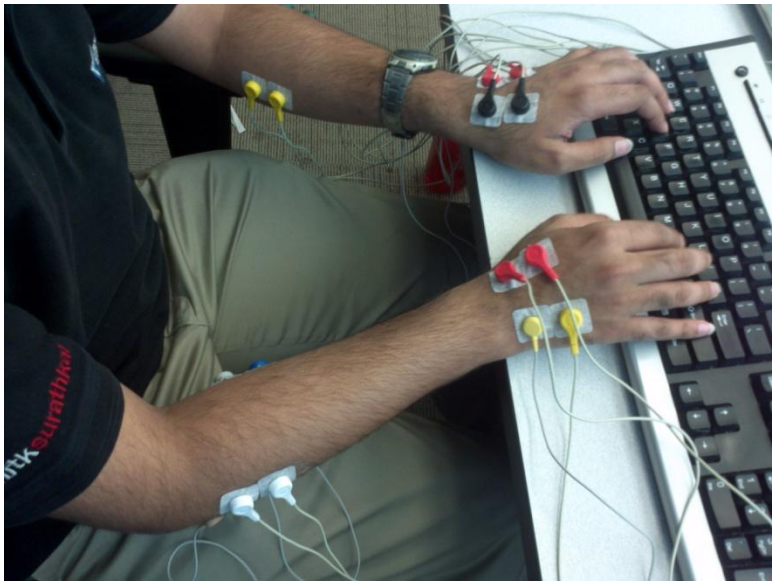

Figure 1 (b). Electrode placements for biometric experiments.

The data was segmented into trials using the keystroke timing data. After manually aligning the start of the first trial in the EMG data with the time recorded at the first keystroke, all trials could quickly be segmented into complete password-length trials. Each trial begins with pressing down the [Left Shift] key, and ends with releasing the [Return] key. Trials can also be checked for errors using the keystroke timing data, and all incorrectly typed trials were discarded.

The trials were of different length, so each trial was divided into M bins. RMS was taken of each bin to both equalize the length of each trial (to length M) and to further smooth the data.

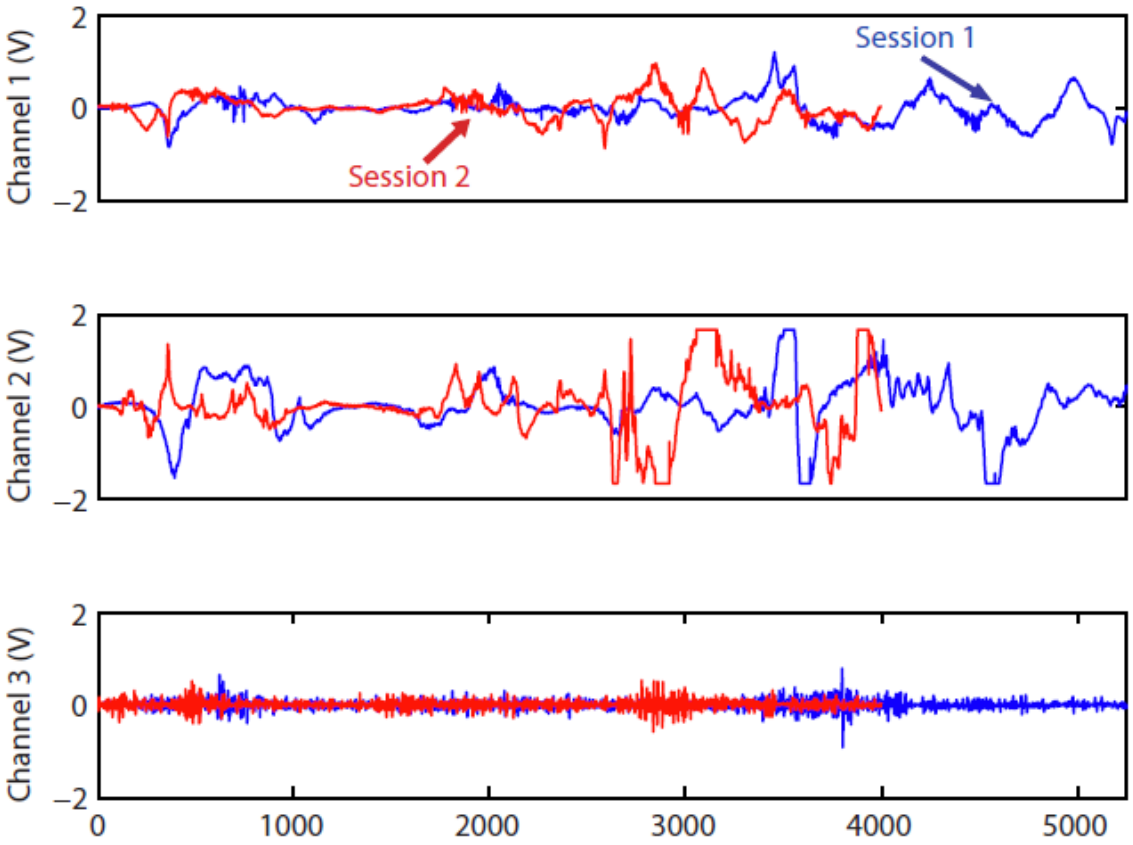

Figure 2 (a).

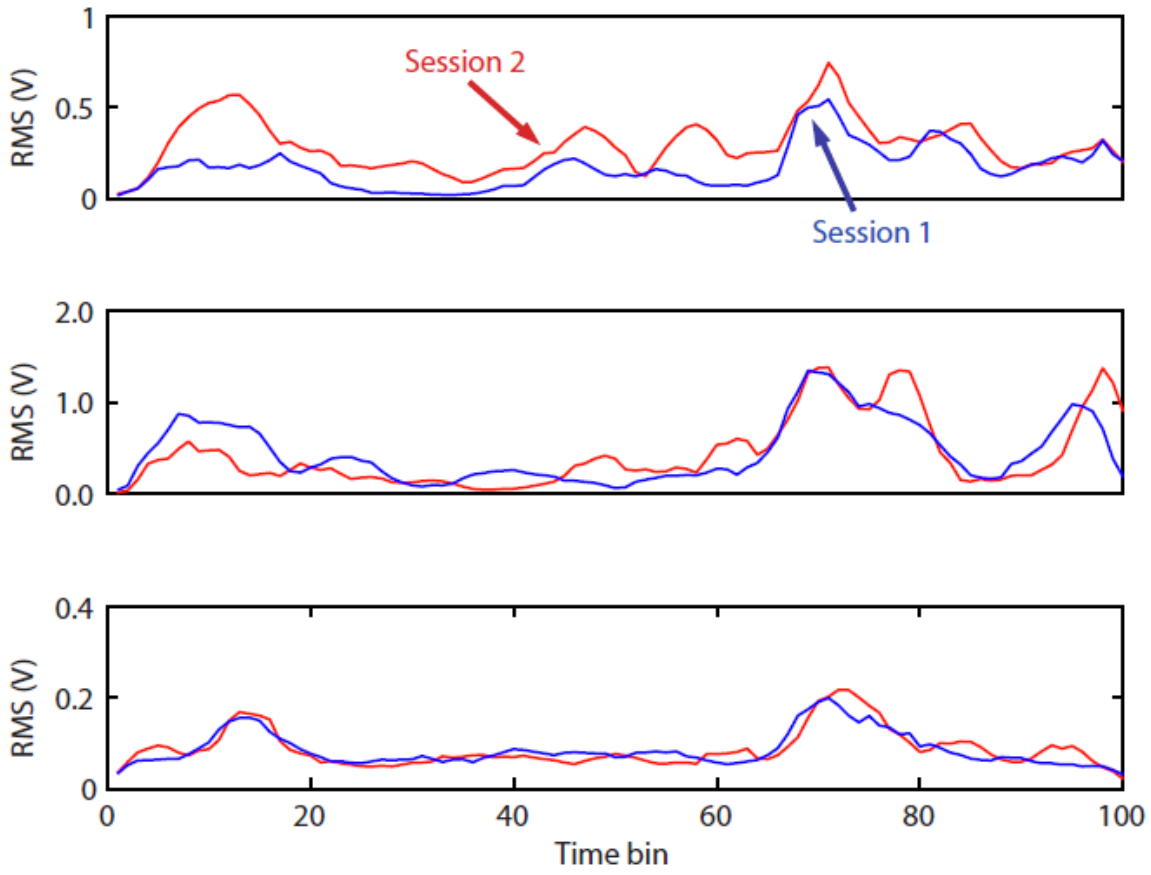

Figure 2 (b). While trials of different length (a) are not apparently similar, taking the RMS of each of M bins produces similar waveforms as seen in (b).

We trained on data collected during session 2 (50 trials) and tested on data collected during session 1 (100 trials), with no cross-validation, because all data in session 2 was used to build Gaussian distributions which were tested with all data in session 1.

Algorithm

We constructed a Gaussian model. Formally, for a given sample $\mathbf{x}_i$, $P(\mathbf{x}_i = \mathbf{x}_i \mid \lambda = \lambda_l) = \mathcal{N}(\mathbf{x} = \mathbf{x}_i \mid \mu_c, \sum_c)$ where $\mathcal{N}(\mu, \sum)$ denotes a Gaussian distribution with mean μ and covariance $\sum$.

Parameters for each class λ_l with K_l samples in session 2 are estimated by the following familiar equations:

$$\mu_l = \frac{1}{K_l} \sum_{i=1}^{K_l} x_i$$

$$\Sigma_l = \frac{1}{K_l}\sum_{i=1}^{K_l}(x_i - \mu_l)(x_i - \mu_l)^T$$

The maximum likelihood estimate of the class label for a given data sample from session 1 is given by:

$$\hat{\lambda} = \underset{l\in\{1,2,\ldots,N\}}{\operatorname{argmin}} \mathcal{N}(x = x_i|\,\mu_l, \Sigma_l)$$

Speed was not an issue, as we were running our operations in MATLAB and performing straightforward operations.

Results

When using Gaussian models, identification rate was 48.68% (random guess is 1/14 = 7.1%). Very low false identification (below 10%) corresponded with verification over 90%.

Posture was a possible source of error. Subjects raised and lowered a chair at their discretion. In future, arm angle relative to the plane of the keyboard should be constant. Fatigue also was a source of error. As the subjects typed roughly 150 trials, focus could drift, which would leave a lot to muscle memory. Inconsistent communication was another factor. For some but not all subjects, we counted off time in between each trial. We also did not have any set script of instructions.

MOTION CLASSIFICATION

Prior Work

Hand, arm, or otherwise, motion classification has been studied by Kator and Lin (S11, G3) in “Super Hand”. They found that while motions could be classified to some degree for a certain subject, a given subject’s motions did not classify well against the other subject’s motions. They pointed to consistency in electrode placement as a possible source of error.

We hope to improve upon classification of motions for a given user, to determine if motions can be classified when training and testing across multiple days, and to better understand the nature of forearm EMG data.

EMG signal classification has been used for a variety of human-computer interactive purposes. Recently, Ju, *et al*, have found ~85% classification results when pressing each individual finger to the thumb by using a hidden Markov model. Saponas, *et al*, demonstrate

average classification accuracies up to 86% for pinching with one of three fingers by incorporating a wireless armband which requires no new calibration or re-training.

Our work is novel due to its experimental simplicity. To our knowledge, free arm motions have not before been studied. We also use a Gaussian model, and use only one feature (trial RMS) per electrode pair. Finally, we allow the subject to essentially define his own motion, as long as it is repeatable.

Data

The CleveMed BioRadio 150 system was used to record four channels of EMG data. One channel consists of two electrodes, where the differential is captured, amplified, filtered, and wirelessly transmitted back to the BioRadio Capture Lite desktop software. Electrode locations are specified by Figure 3. Filters used were identical to those described in the biometrics section of this report.

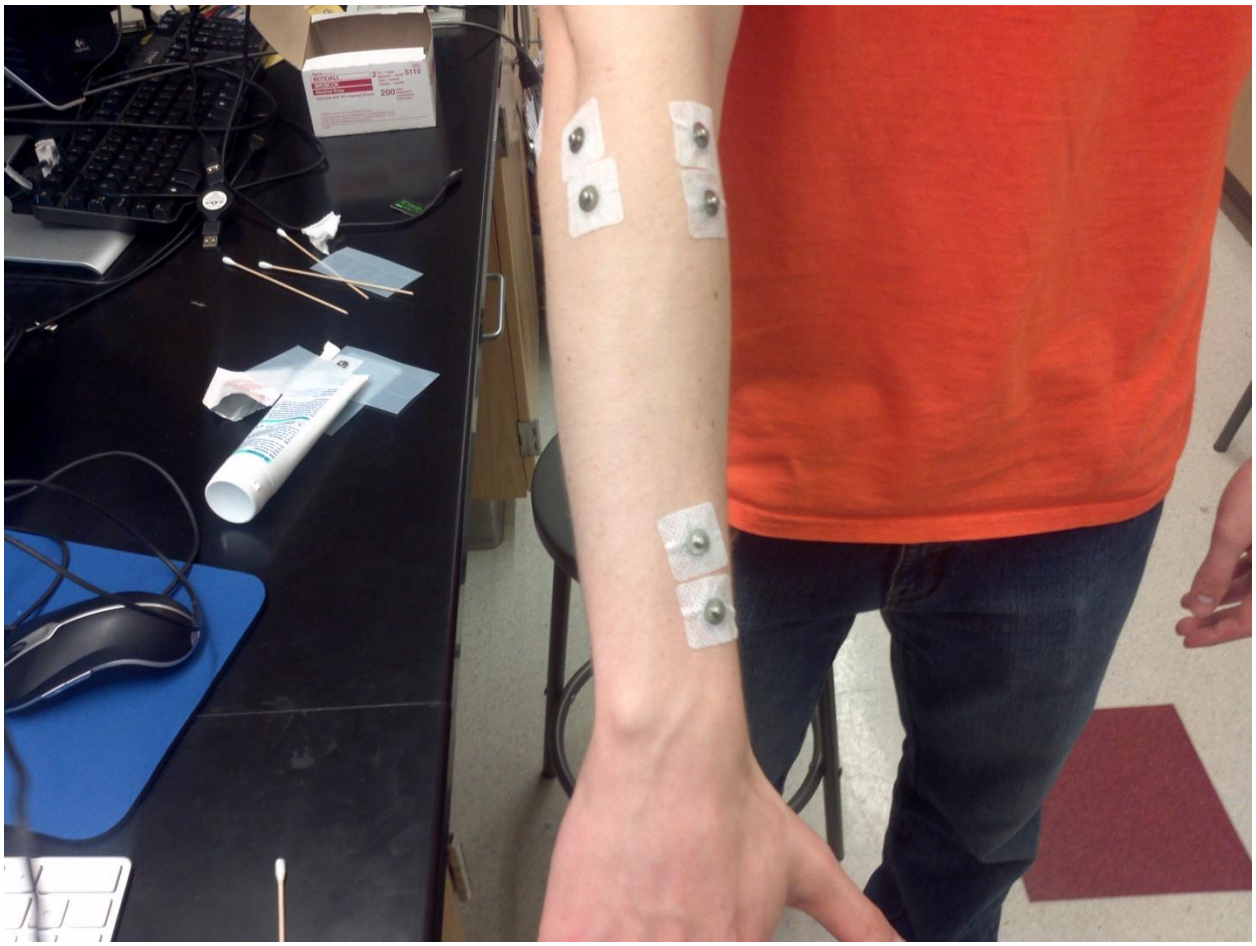

Figure 3 (a)

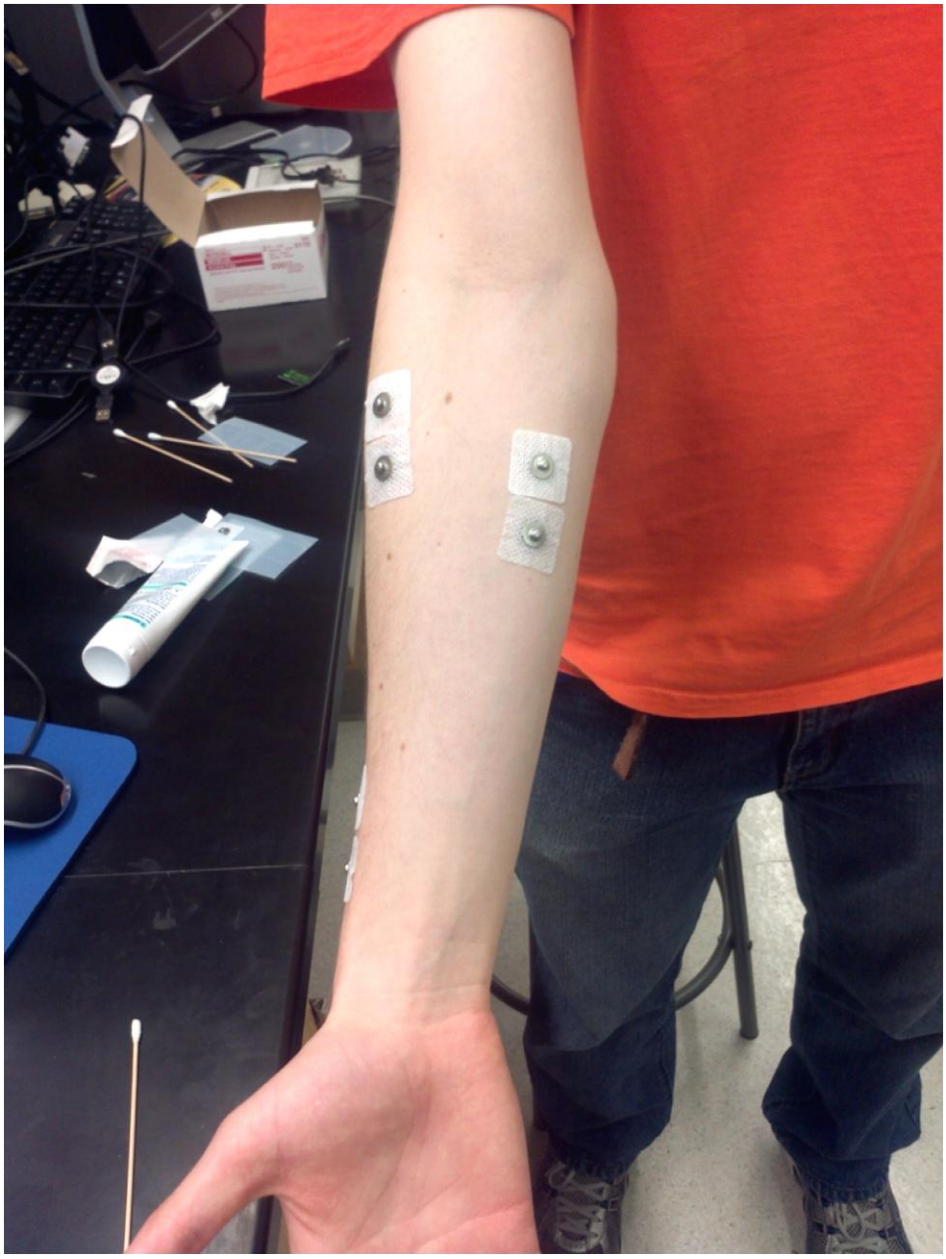

Figure 3 (b).

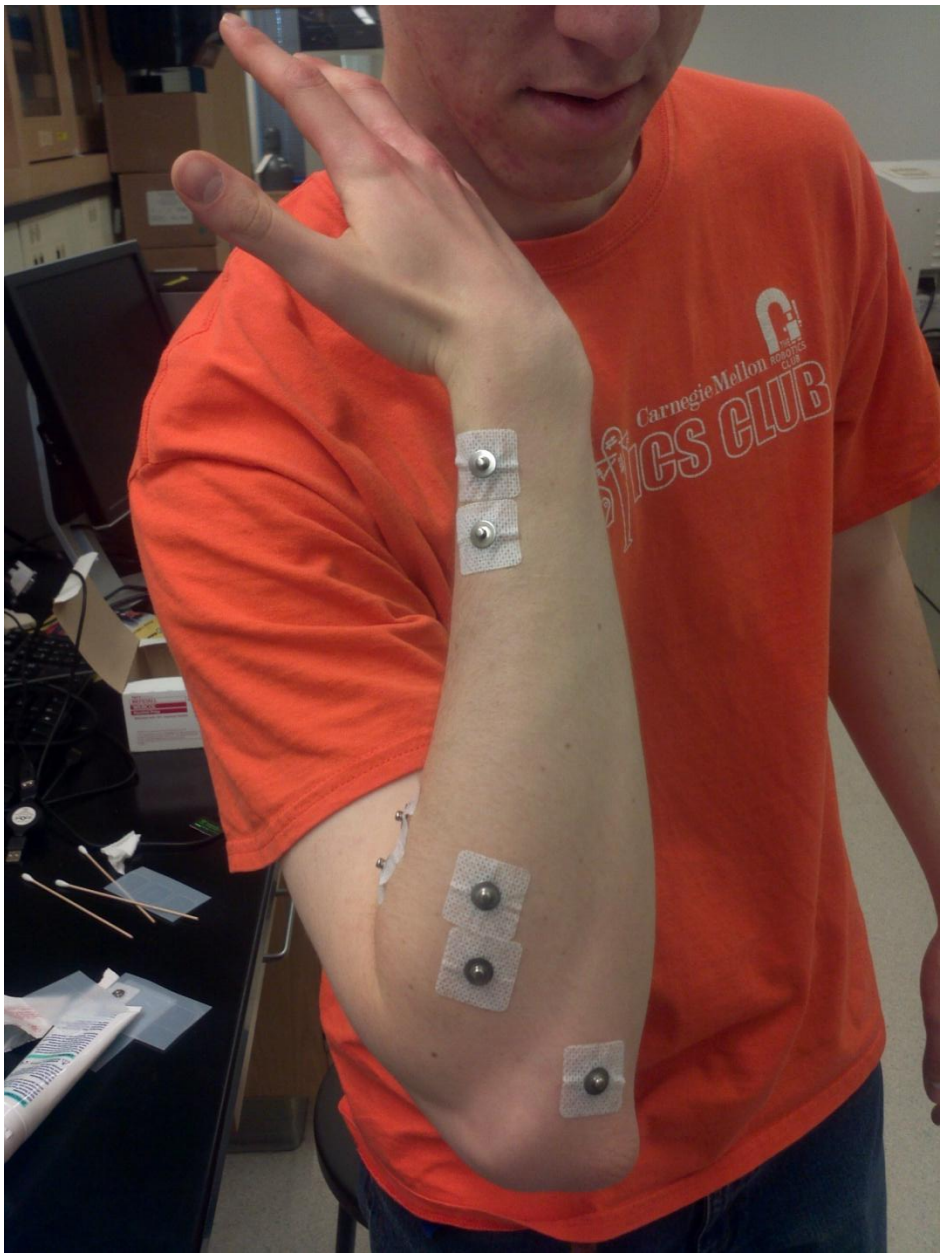

Figure 3 (c). Electrode placement for motion classification.

Segmentation was an important first step when obtaining the trial data. For accurate classification, we required many datapoints from each user. For example, the total number of datapoints collected in all our experiments was over 5600 trials. This amount would not have been possible without automatic segmentation of trials.

For one session, the subject had to make repeated arm movements (denoting trials), with inactivity between trials. One naïve solution would save the data after every trial, severely constraining the ease and time it takes to collect data. However, we saw the need to

develop an algorithm that would allow the user to continuously repeat trials, significantly cutting down the overhead time in experiments.

The basic idea of the algorithm is to find the max peaks of the EMG's RMS for each channel. By labeling the peaks, we can take a small chunk (one second) as a trial, while disregarding the intervals between trials. To find the max peaks, we first use a sliding window on a channel's signal. At each time step, the maximum peak time is recorded in a vector, which can include all channels' peak times. After obtaining the vector, we use non-overlapping bins (with a user-defined length) to find the average max peak time within each bin. This ensures that the most dominant maximum peak times (i.e. as the window was sliding, it kept returning the same peak time) are represented. This method is prone to extraneous trials (e.g. false positive max-peaks), and a visualization component in which extraneous trials may be deselected was implemented.

Segmentation Algorithm:

1. Compute the root-mean-square with convolution
2. Find the maximum value for each step of the sliding window
3. Average the max values in non-overlapping bins over selected channels
4. Manually review the selected trials, deselecting the extraneous ones
5. Capture 1 second of data from each selected max peak time

Included are the visualizations of the segmentations (without deselected trials). The user would need to identify the max-peak times (with the corresponding numbers on the x-axis) that do not represent a trial.

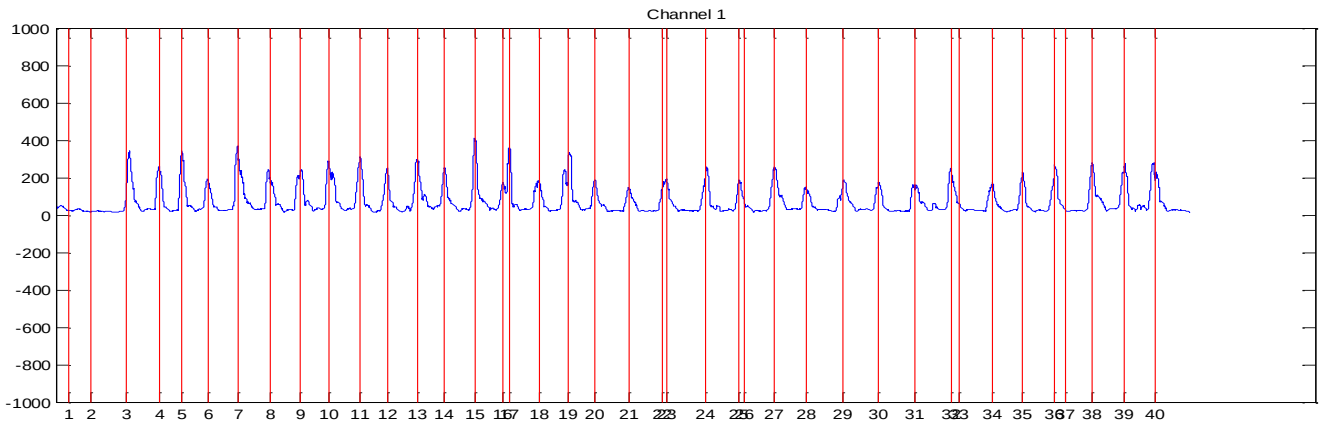

Figure 4. Non-manually curated segmentation.

Algorithms

Data collection was obtained using the CleveLab's software. All algorithms, including filtering, segmentation, and classification, were personally written in Matlab. We did not use any third-party software for that portion. We did not encounter performance issues. All code was included in the final project.

The segmentation code's parameters must be modified by the user to perform proper segmentation. Extraneous max-times must also be removed by the user.

The classification code consists of two parts: cross-validation and the actual classifier. The cross-validation algorithm calls the classifier code.

Hand Motion Classification

Hand motion classification is one of the largest areas of active research on electromyography. The literature has shown that simultaneously recording from multiple muscle groups has yielded high motion classification accuracies [1, 2]. These promising results provide incentive to design and implement next-generation prosthetic devices that are controlled by EMG data from the user. However, many questions still need to be answered, such as the number of necessary muscle groups, the amount of noise in the signal, and the lifespan of a static decoder without retraining.

Motivation

Leveraging our results from our biometric experimentation, we explored how EMG activity varied over multiple sessions. While most studies have relied on trials collected in a short time span [2], we tested how the time between trials affected classification results. These results not only characterize the stability of an EMG system but also the robustness against external noise in the signal. The ability for a decoder to provide accurate classifications without the need for retraining is a highly desirable quality, as a prosthetic device could remain operational for months without the patient spending a significant amount of time re-calibrating the system. More importantly, if a probabilistic model can describe the external noise in the system (such as disturbance from electrode shift or decay), more appropriate dynamic decoders may be used to account for these effects.

Visualization

For visualization purposes, principal component analysis (PCA) was applied to each channel's waveform to find the underlying eigenvector waveforms. To achieve this for a channel, each timepoint was considered a dimension (with 960 timepoints per trial). PCA was then performed on each task, concatenated over sessions. This analysis reveals that

each task has characteristic waveforms across channels, which encouraged further classification results. In figure 3, tasks are denoted by color.

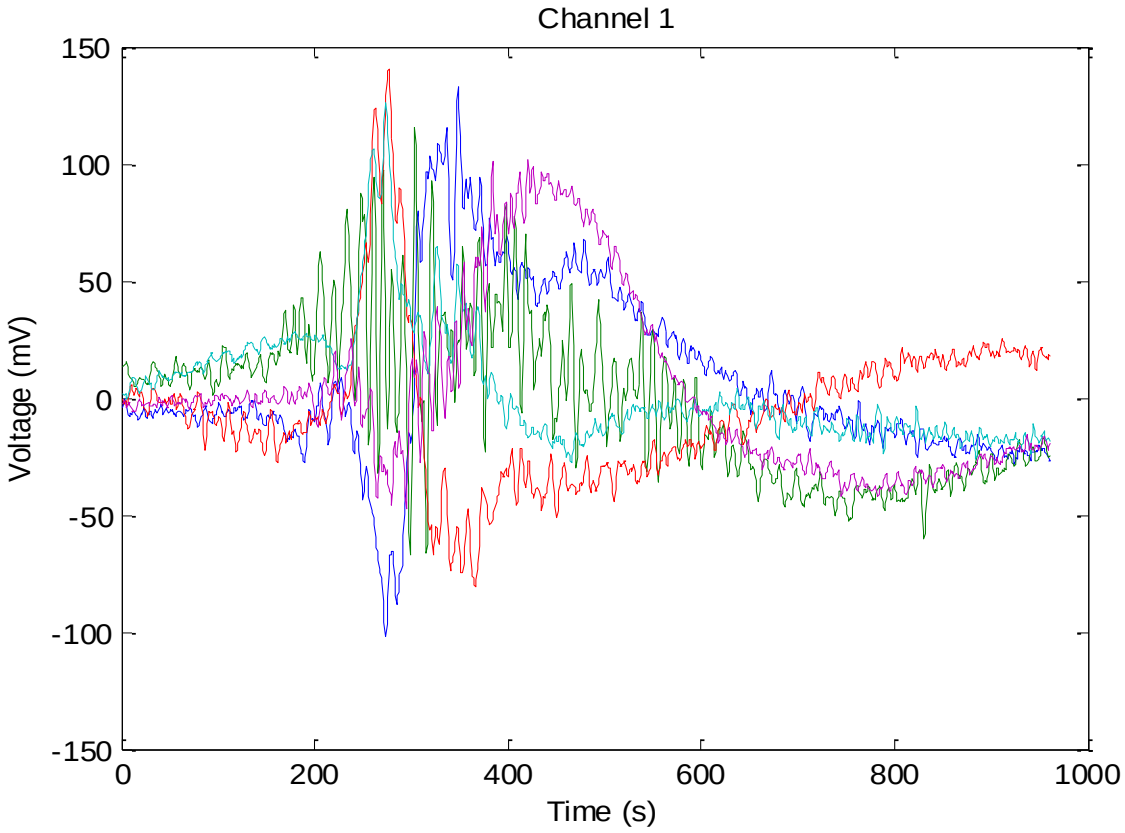

Figure 5 (a)

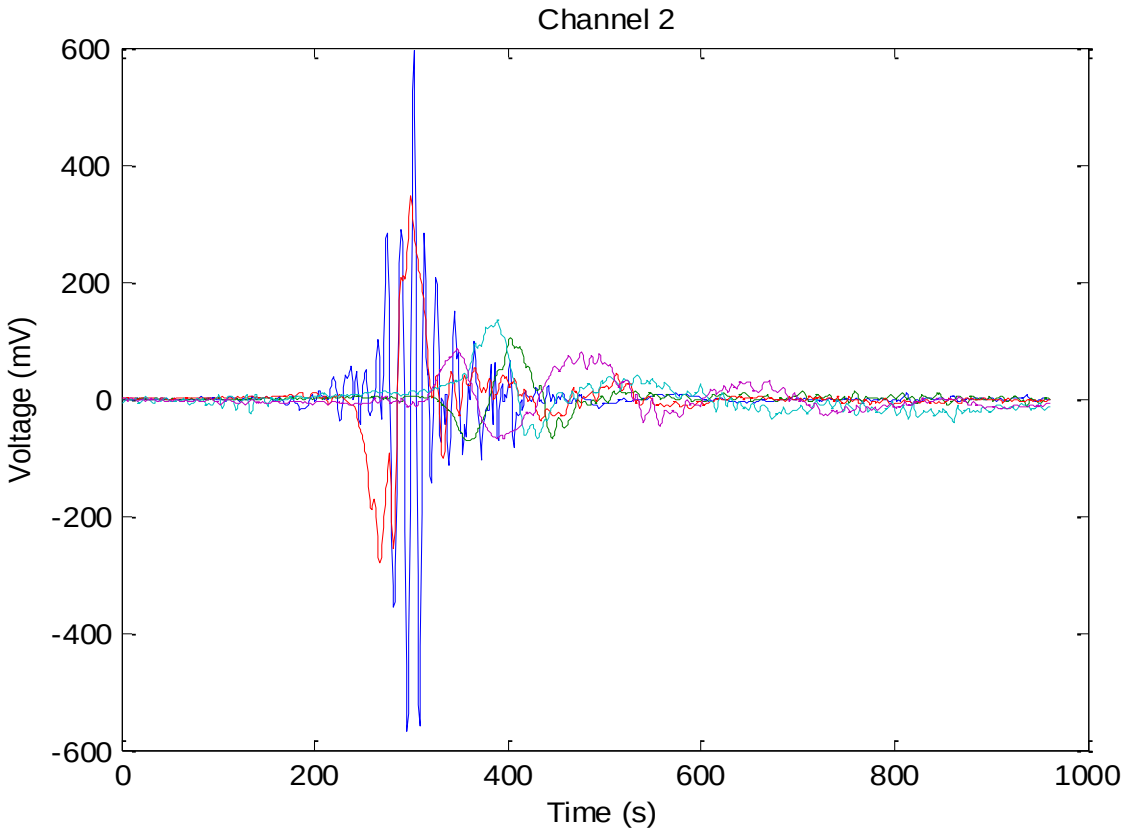
Channel 2
600
400
200
0
-200
-400
-600
Voltage (mV)
0
200
400
600
800
1000
Time (s)

Figure 5 (b)

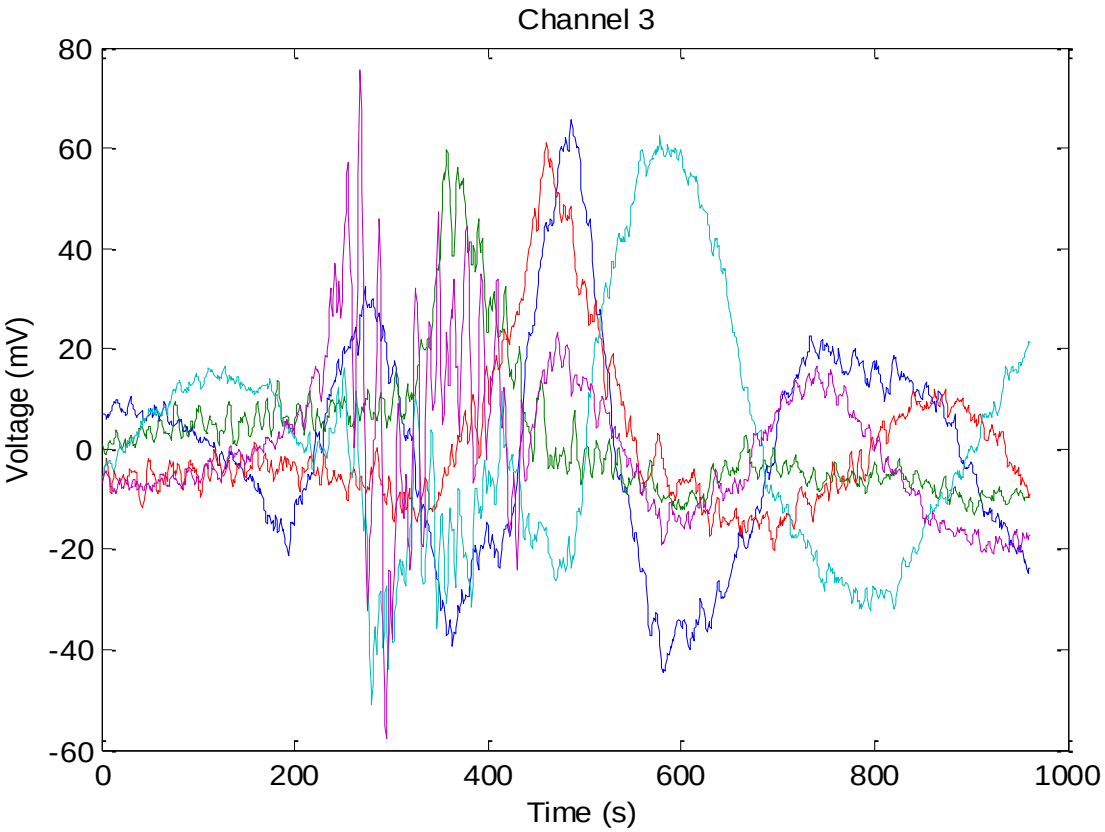
Channel 3
Voltage (mV)
Time (s)
80
60
40
20
0
-20
-40
-60
0
200
400
600
800
1000

Figure 5 (c)

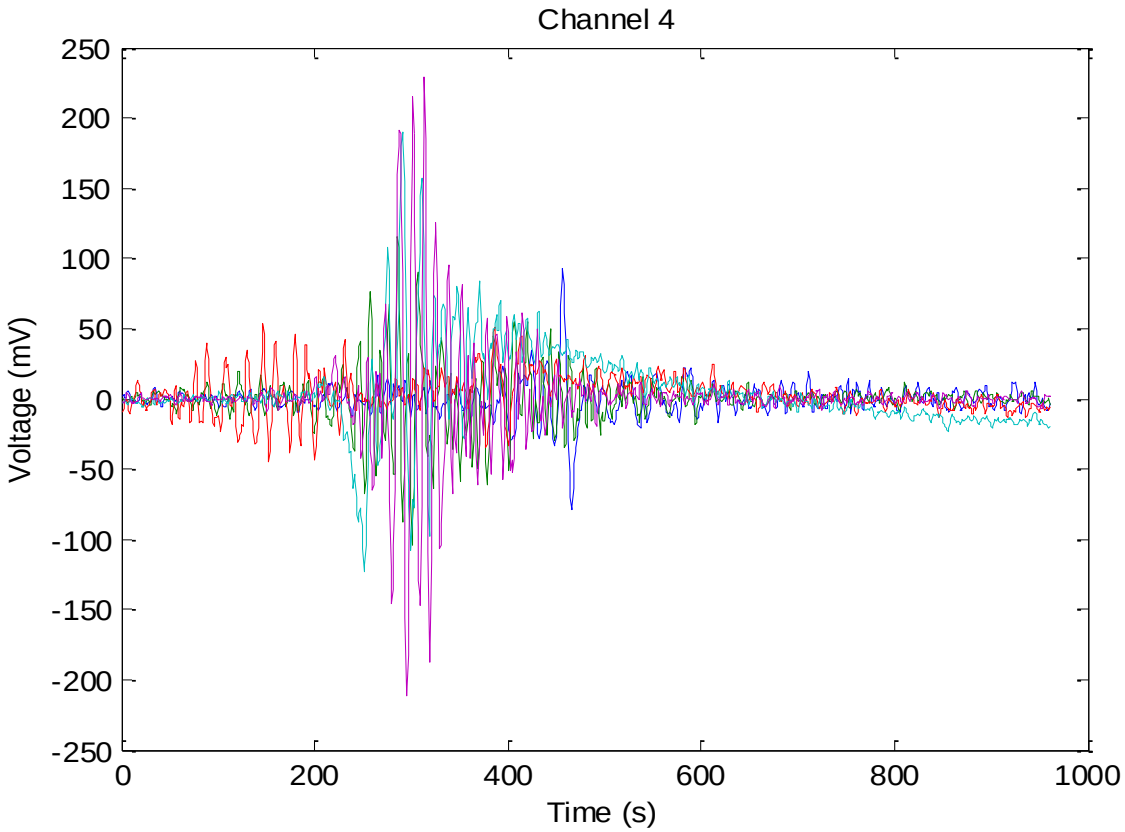

Figure 5 (d). Figure 5 shows principal components for each motion.

Also, PCA was performed over the entire space to find characteristic waveforms of each muscle. These waveforms can form the basis for PC Smoothing, a method in which the raw signal is projected into a lower-d waveform space, and then projected back into the original space. The result is a denoised waveform that has extracted the variance from the most important dimensions. However, this method may not capture the most meaningful dimensions of variance.

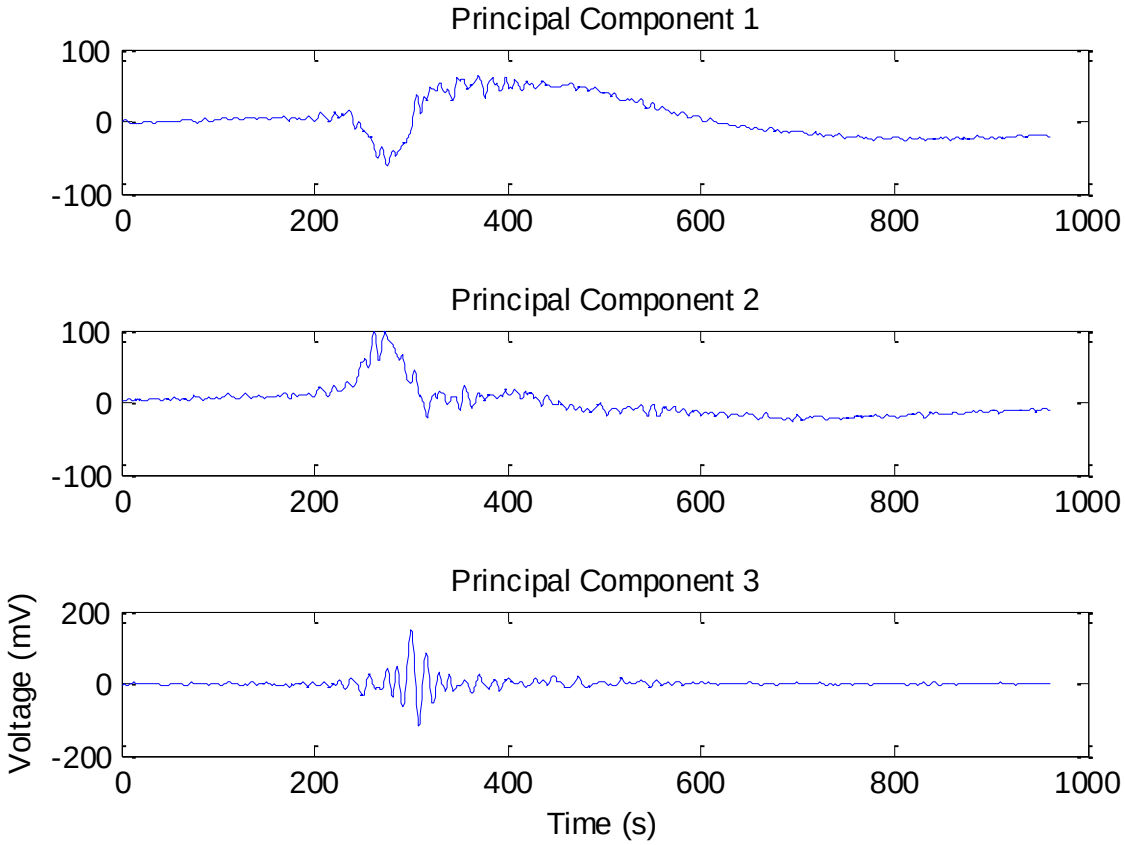

Figure 6. Principal components over the entire channel and motion space may reveal muscle activities.

We preferred root-mean-square (RMS) over PC Smoothing, as the former makes less assumptions about the data and significantly reduces the number of dimensions. The RMS of the raw EMG signal was taken for the entire trial, producing a 4x1 vector where each element corresponds to a channel. This implies each trial can be represented in a four-dimensional space. We used DataHigh, a novel data visualization tool to view many projections of the 4d space. Each point corresponds to one trial, and each color corresponds to one task. As one may see (FIGURE), the clusters are well separated, and assume a Gaussian cloud shape. This motivated which classifier we ultimately selected.

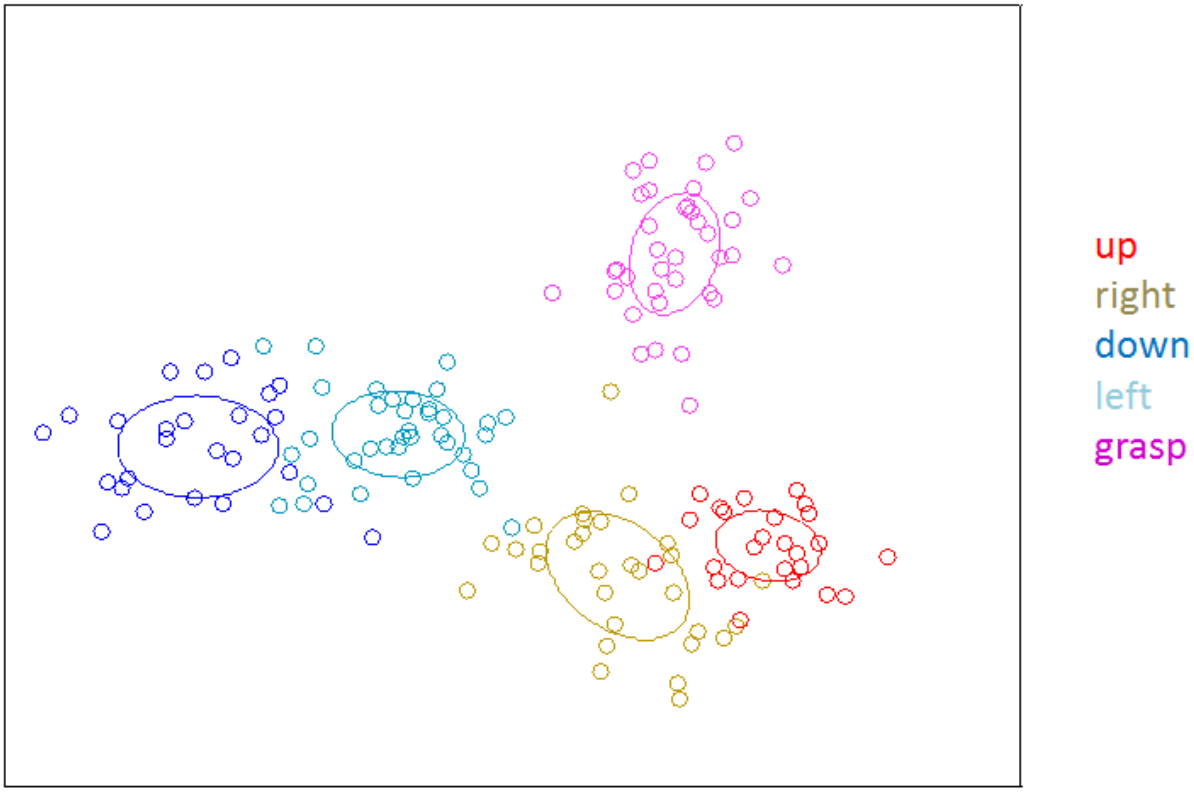

Figure 7 (a).

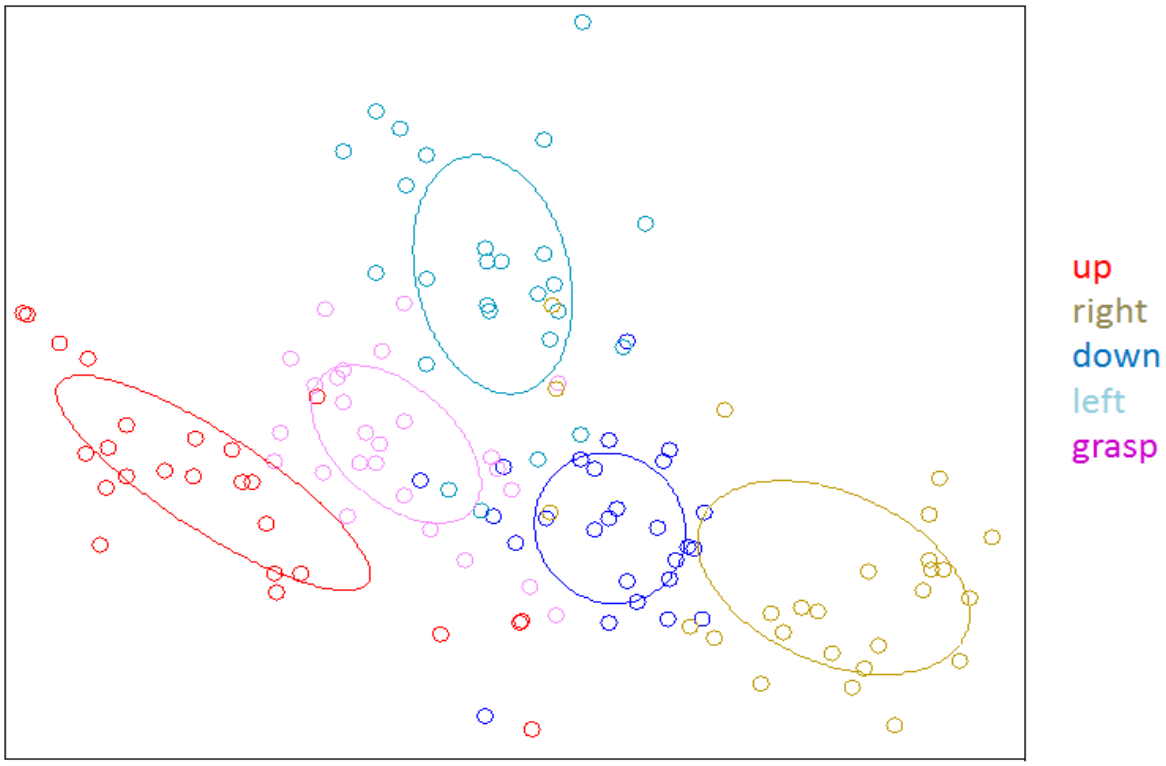

Figure 7 (b). Figure 7 shows 2d projections of the 4d RMS space. Subject Cow, Session 8 shown in (a), Subject Shim, Session 8 shown in (b). Ellipses show one standard deviation of a Gaussian covariance matrix.

Method of Classification

After visualizing the data, we decided to use a Gaussian Maximum-Likelihood Estimator (MLE) that would be well suited for the Gaussian clusters. Given the training data, the MLE decoder fits a Gaussian to each class. Then, for each test instance, the Gaussian yielding the highest probability is chosen. This probabilistic decoder has several advantages, providing a generative model and a Bayesian framework for decoding. However, if there are not enough trials, over-fitting is a serious issue. More importantly, if the data does not conform to a Gaussian distribution, the MLE decoder's performance could rapidly decline. To prevent over-fitting, factor analysis may be used to better approximate the covariance matrices ($\Sigma = LL' + \varphi$). In this case, we had to fit 4x4 covariance matrices. Thus, while over-fitting could be an issue, we felt that 25 or more trials per class would provide a good fit for each Gaussian without using factor analysis.

Results for Within-Session Classification

Using 5-fold cross-validation, we confirmed high accuracy for within-class classification. For one fold, the classifier is trained on 100 training samples (20 samples from each class), and

then tested with the remaining 25 samples. This process was repeated for all eleven sessions. Results for both subjects are high (Figure 6), with each session having greater than 70% correctly classified test labels. There are no significant trends between sessions.

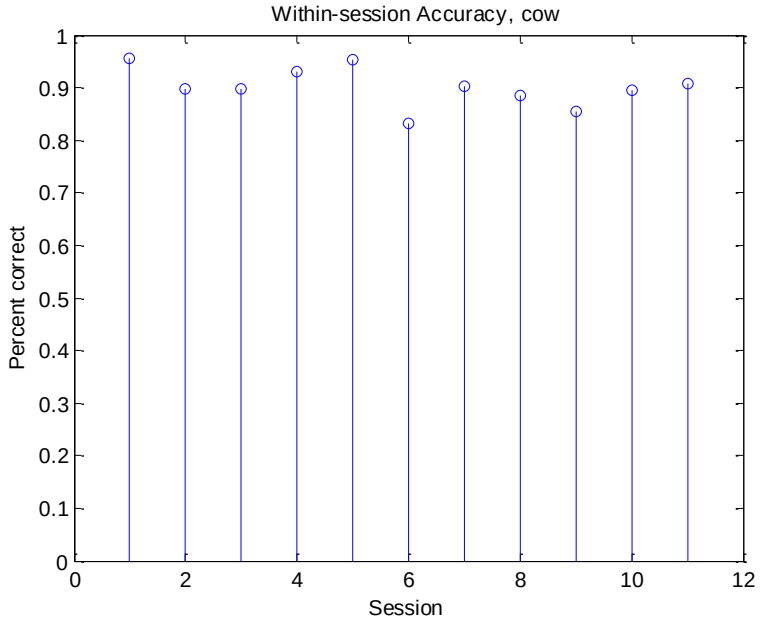

Figure 8 (a).

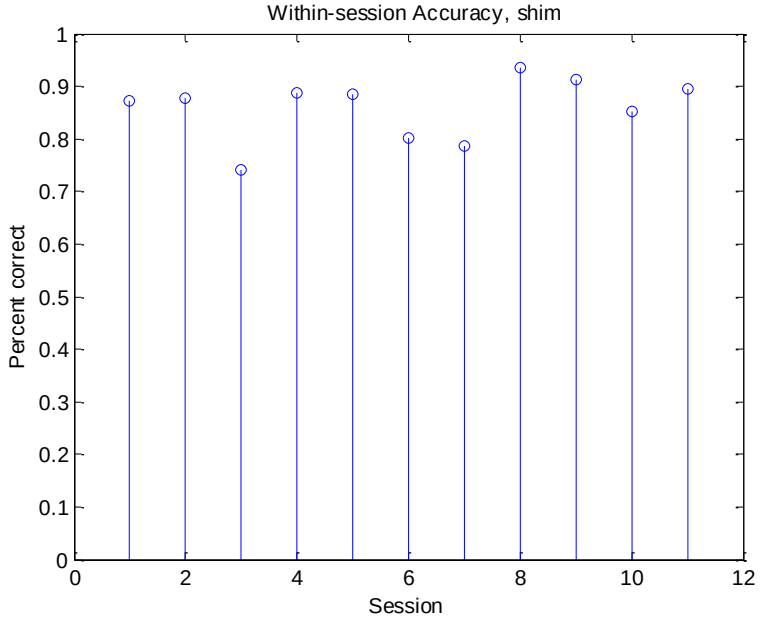

Figure 8 (b).

One may further analyze the confusion matrices for particular sessions. Shown (Figure 9) are the confusion matrices for two different sessions of one subject. The vertical axis represents the actual class, and the horizontal axis denotes which class the decoder selected. For this particular subject, his **down** and **left** movements were very similar, while his **up** and **right** movements also were difficult to differentiate. His grasp motion was easily classified against the rest.

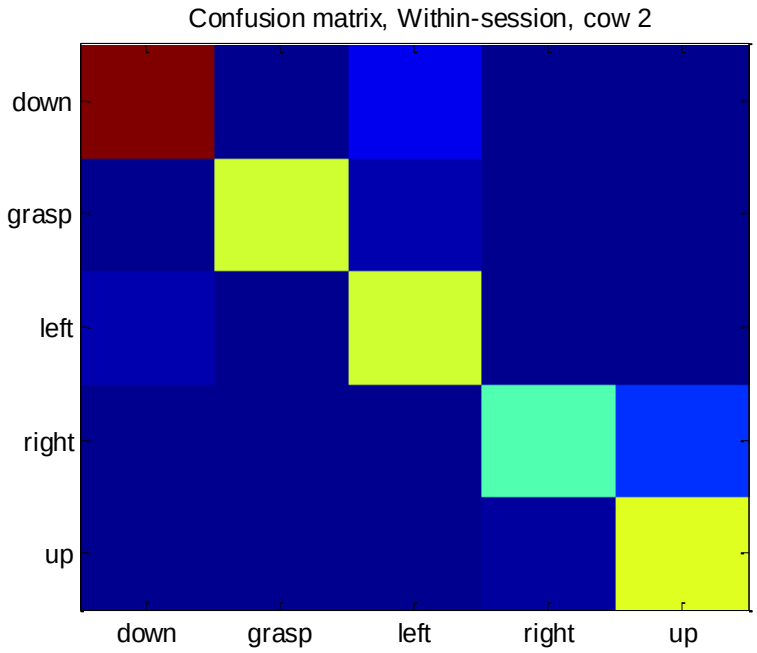

Figure 9 (a).

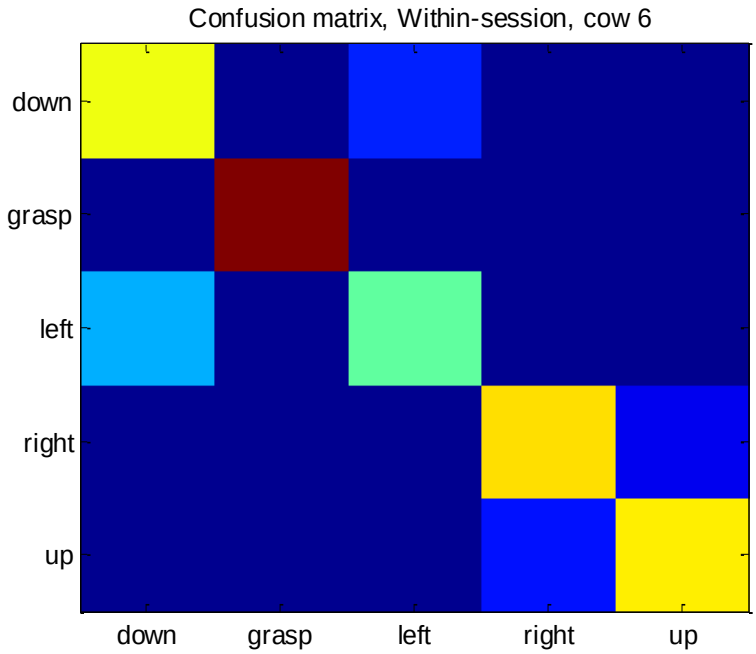

Figure 9 (b).

Encouraged by our within-session classification results, we proceeded to test between-session classification. To achieve this, we trained on the data from one session (typically 125 trials) and tested on another session (again, ~125 trials). To be thorough, we trained on every session, and tested that decoder with every other session. We generated a matrix to show between-session classification accuracy (Figure 10). Across the rows shows the classification accuracy for a trained session. Sessions 1 through 6 were carried out on the first day (each separated by at least a 15-minute interval). Sessions 7 through 11 were carried out on the preceding day (the electrodes were not replaced). Classification accuracies were higher within single days than across days, which should be expected. Interestingly, results were higher when trained on a Day 1 Session and tested on a Day 2 session than the other way around. These results show that EMG can be stable across sessions within a single day. However, future work will need to explore how the EMG data changes across days (Figure 11).

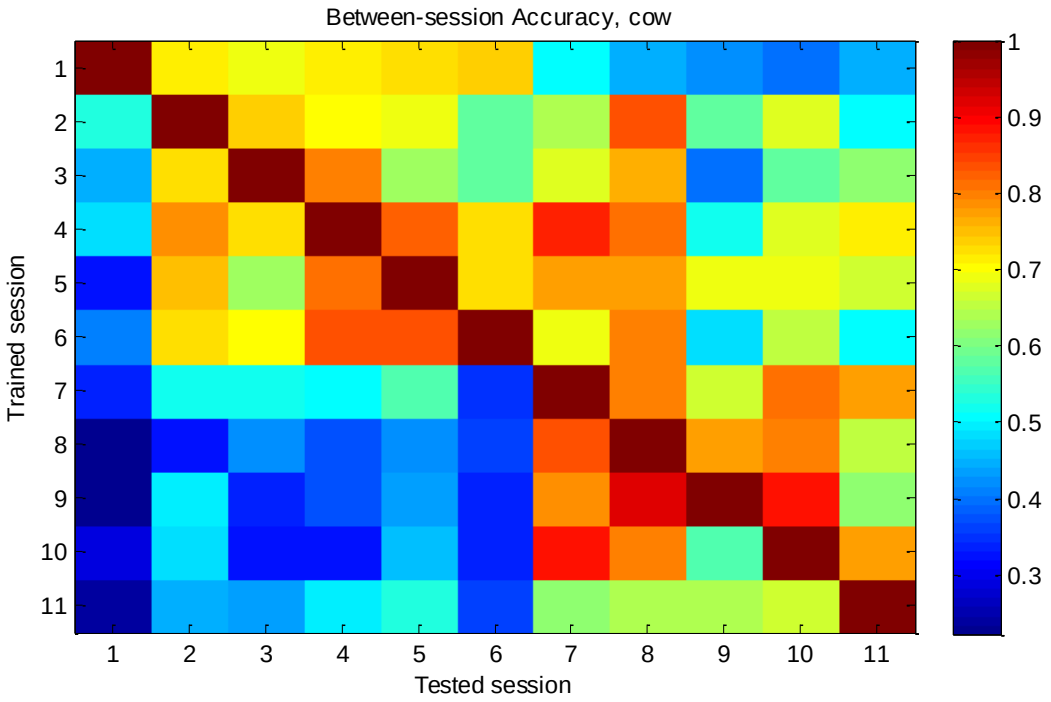

Figure 10 (a).

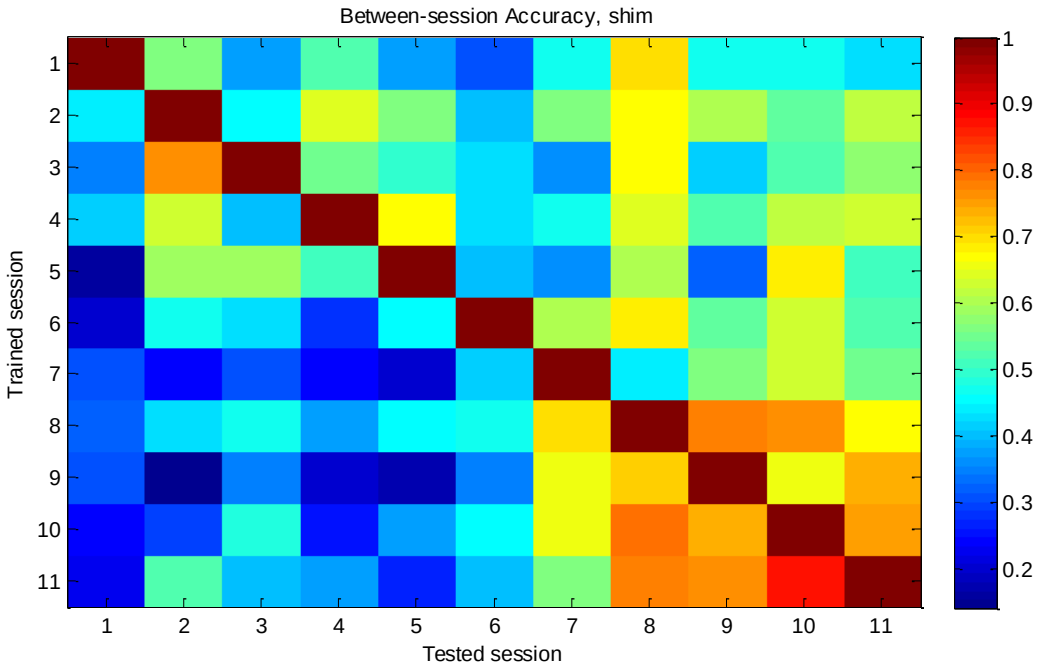

Figure 10 (b).

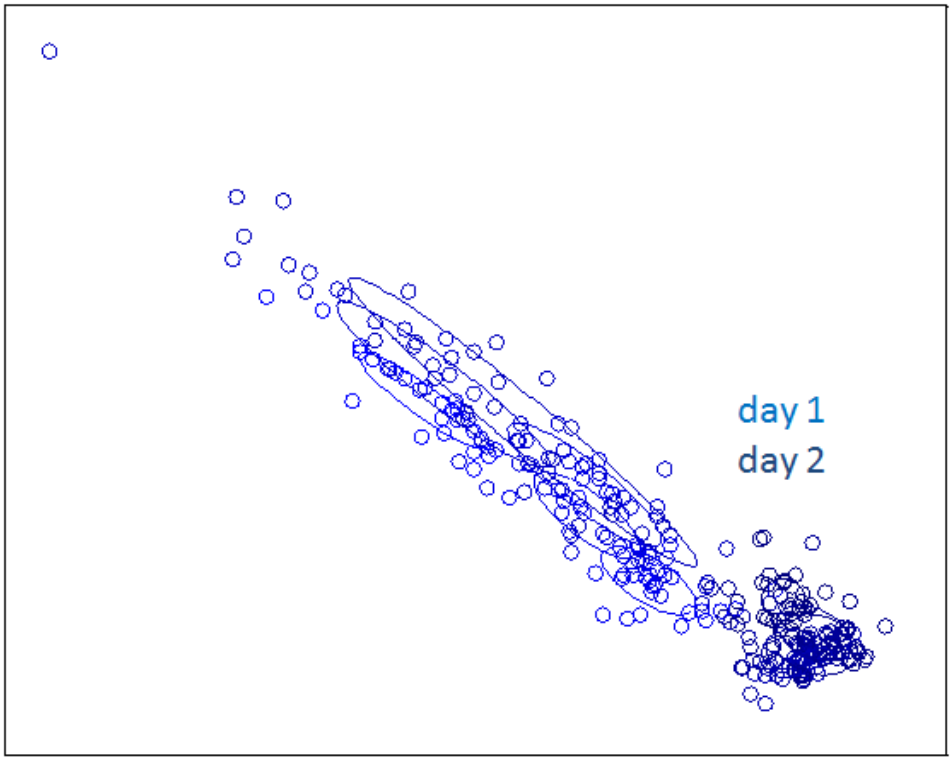

Figure 11. Differences between sessions when over 24 hours apart.

In Figure 11, the 2d projection of the 4-d space shows the progression of one channel's clusters over the course of eleven sessions. Each covariance ellipse signifies one session. Day 1 has greater variance (center Gaussian clusters) as opposed to Day 2 (bottom right Gaussian clusters).

We also wanted to see the results if we did 10-fold cross-validation on all sessions. Explicitly, we trained on ten sessions and tested on the remaining session. Results are shown for both subjects (Figure 12). By including sessions over two days, the data is no longer a strong Gaussian fit, and errors ensue. The training set size was 1250 samples, while the testing set size was 125.

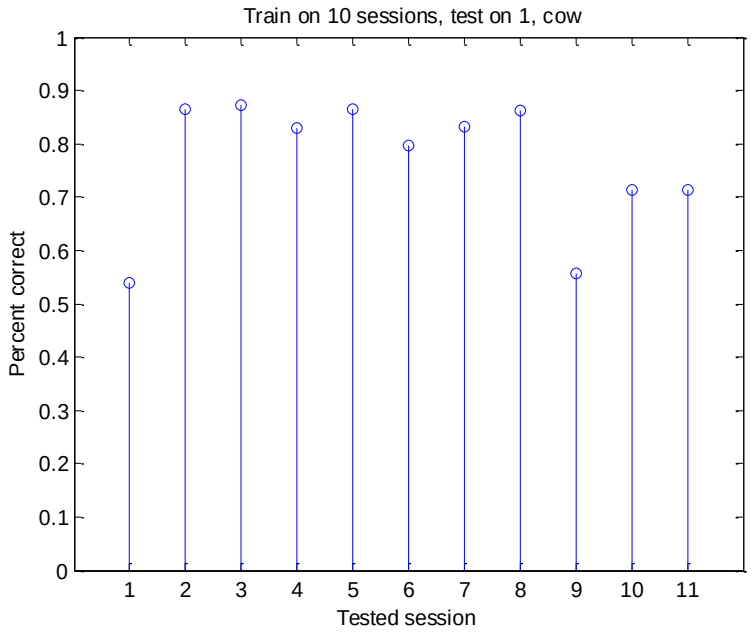

Figure 12 (a)

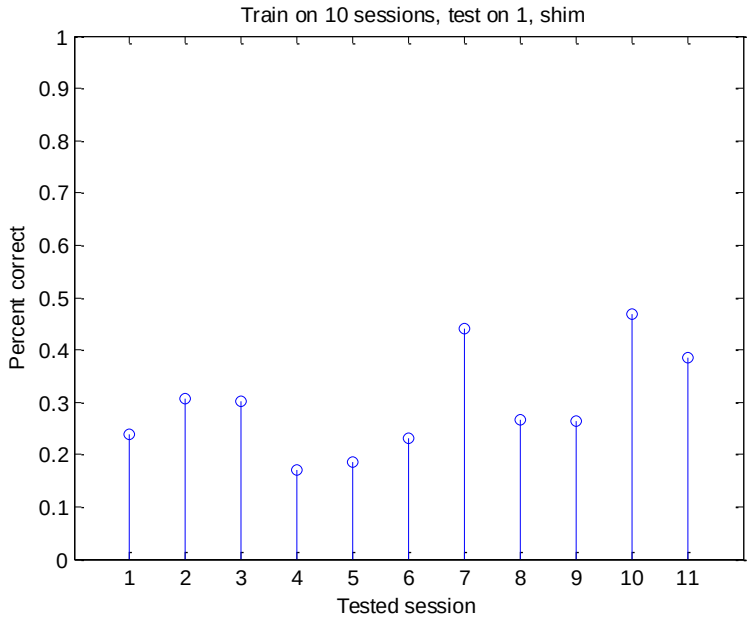

Figure 12 (b). Subject Shim had generally inconsistent hand motions, while subject Cow was more consistent. Random guess is 20%.

To buttress our previous hypothesis about biometric testing, we decided to test biometric accuracy on hand motion data. With two subjects, chance is 50%. We trained on 20 sessions (i.e. 10 sessions from each subject), and had the task to classify trials from two sessions (each session from a different subject). Thus, we had a total of 2500 training

samples and 250 test samples. As shown, our accuracies were extremely high. Intuitively, free-form hand motion classification would warrant high differentiability between subjects, as each subject may have a unique sequence of motor actions to produce a movement. For example, Subject Cow used mostly wrist motion, while Subject Shim included significant finger motion.

Tested Sessions	Down	Grasp	Left	Right	Up
1	1.00	1.00	1.00	1.00	1.00
2	1.00	1.00	1.00	1.00	1.00
3	1.00	0.98	1.00	1.00	1.00
4	1.00	1.00	1.00	1.00	1.00
5	1.00	1.00	1.00	1.00	1.00
6	1.00	0.95	1.00	0.86	1.00
7	1.00	0.98	0.98	0.88	1.00
8	1.00	0.97	0.98	0.94	1.00
9	1.00	0.96	0.98	0.96	1.00
10	1.00	0.97	0.94	0.96	0.96
11	0.98	0.98	1.00	1.00	0.98

Figure 13. Biometric feasibility of personal user hand motions. Random guess is 0.50.

Demo

Our demo consisted of taking a short session of data (“grasp” motion) from subject Shim, training a classifier based on this new data combined with a single session from day 2 for all the other motions, and generating a confusion matrix. We demonstrated a diagonal confusion matrix, even though the grasp motion was trained and tested on data from our demo, and the other four motions were trained and tested on data from a much earlier day, even with electrode placement not being identical to previous days.

Debriefing

Schedule

Task	Responsible
Background Research	Shimomura

Algorithm Design	Both
Data Collection	Both
Algorithms Coding	Cowley
Presentations and Reports	Both

In future, we would like to add more electrodes, possibly incorporated into a textile. Increasing the number of electrodes would then make machine learning techniques like factor analysis as a useful option, as the latent variables would allow for better classification. Armbands would also more rigidly enforce electrode placement, rendering possibly confounding variables moot. If electrode placement is still an issue, we recommend also placing an electrode on the brachio radialus, as we discovered this muscle group could be a useful source of EMG data. More users for motion classification would be ideal, but possibly unlikely due to time constraints. Finally, real-time or pseudo-real-time classification would be another step toward Human-Computer Interfaces.

Defining an experimental procedure should be a high initial priority. Posture, defining the extent of “natural movements” (e.g. how much the user is allowed to define his own motion), and the ability to repeat motions are all important experimental points that we made up as we went along.

We have great thanks for the entire lab of Marios Savvides, PhD, and the help of Conrad Zapanta, PhD during our entire process.

REFERENCES

1. Ju P, Kaelbling L, and Singer Y. “State-based classification of finger gestures from electromyographic signals.” Presented at the ICML Stanford, CA, 2000.
2. Saponis TS, et al. “Making Muscle-Computer Interfaces More Practical.” Proceedings of CHI 2010.
3. Suresh M, Krishnamohan PG, Holi MS. “GMM modeling of person information from EMG signals.” Recent Advances in Intelligent Computational Systems (RAICS). 2011 IEEE, 712-717, 22-24 Sept. 2011.
4. Ahsan MR, Ibrahimy MI, Khalifa OO. “EMG signal classification for human computer interaction: a review.” 2009. European Journal of Scientific Research, 33 (3). pp. 480-501.
5. R. Merletti, P. Parker. *Electromyography – Physiology, Engineering and Noninvasive Applications* (first ed.). John Wiley & Sons, Inc., Hoboken, New Jersey (2004).
6. R. Merletti, Standards for reporting EMG data, J. Electromyogr. Kinesiol. 9 (1999) III–IV.

7. D. Farina, T. Lorrain, F. Negro, N. Jiang. "High-density EMG E-Textile systems for the control of active prostheses," Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE , pp. 3591-3593. Aug. 31 2010-Sept. 4 2010
8. Kator A and Lin M. “Super Hand.” Presented for 18-551 Design, Spring 2011.
9. Jordan M, Hsieh PH, and Sanghvi S. “Find MY FACE!” Presented for 18-551 Design, Fall 2009.
10. Devi A and Seshadri S. “Identity Recognition.” Presented for 18-551 Design, Spring 2011.