Demonstrating the Feasibility of Using Forearm Electromyography for Muscle-Computer Interfaces

T. Scott Saponas[1], Desney S. Tan[2], Dan Morris[2], Ravin Balakrishnan[3]

[1]DUB Group, CS and Eng Dept
University of Washington
Seattle, WA (USA)
ssaponas@cs.washington.edu

[2]Microsoft Research
Redmond, WA (USA)
{desney, dan}@microsoft.com

[3]Department of Computer Science
University of Toronto
Toronto, Ontario (Canada)
ravin@dgp.toronto.edu

ABSTRACT

We explore the feasibility of muscle-computer interfaces (muCIs): an interaction methodology that directly senses and decodes human muscular activity rather than relying on physical device actuation or user actions that are externally visible or audible. As a first step towards realizing the muCI concept, we conducted an experiment to explore the potential of exploiting muscular sensing and processing technologies for muCIs. We present results demonstrating accurate gesture classification with an off-the-shelf electromyography (EMG) device. Specifically, using 10 sensors worn in a narrow band around the upper forearm, we were able to differentiate position and pressure of finger presses, as well as classify tapping and lifting gestures across all five fingers. We conclude with discussion of the implications of our results for future muCI designs.

Author Keywords: Electromyography (EMG), Muscle-Computer Interfaces (muCIs), input, interaction.

ACM Classification Keywords: H.1.2 [User/Machine Systems]; H.5.2 [User Interfaces]: Input devices and strategies; B.4.2 [Input/Output Devices]: Channels and controllers

INTRODUCTION

Many human-computer interaction technologies are currently mediated by physical transducers such as mice, keyboards, pens, dials, and touch-sensitive surfaces. While these transducers have enabled powerful interaction paradigms and leverage our human expertise in interacting with physical objects, they tether computation to a physical artifact that has to be within reach of the user.

As computing and displays begin to integrate more seamlessly into our environment and are used in situations where the user is not always focused on the computing task, it is important to consider mechanisms for acquiring human input that may not necessarily require direct manipulation of a physical implement. For example, drivers attempting to query their vehicle navigation systems may find it advantageous to be able to do so without removing their hands from the steering wheel, and a person in a meeting may want to unobtrusively communicate with someone outside. Also, since physical computer input devices have been shown to be prone to collecting microbial contamination in sterile environments [7], techniques that alleviate the need for these implements could be useful in surgical and cleanroom settings.

To date, most efforts at enabling implement-free interaction have focused on speech and computer vision, both of which have made significant strides in recent years but remain prone to interference from environmental noise and require that the user make motions or sounds that can be sensed externally and by definition cannot be easily concealed from those around them.

Advances in muscular sensing and processing technologies provide us with the unprecedented opportunity to potentially interface directly with human muscle activity. Electromyography (EMG), which measures electrical potentials generated by muscle cells, is one such technology. Driven by the growing societal recognition of the needs of people with physical disabilities, researchers are using EMG to diagnose muscular diseases, assess progress in muscular rehabilitation, and directly control prosthetics [1,8,9,11,18]. In these systems, researchers use detailed knowledge of the
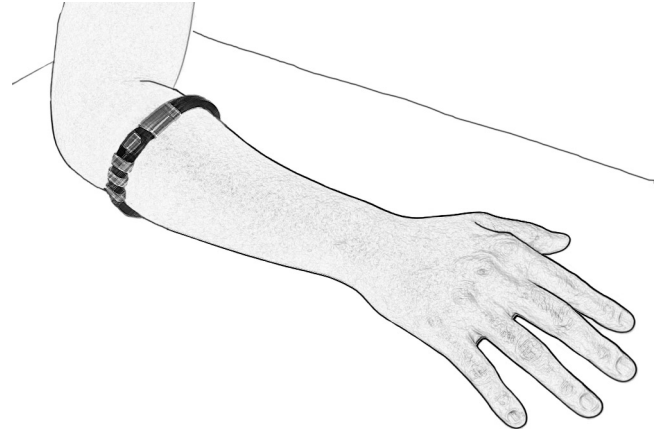
Figure 1: Artist rendering of a forearm band with EMG sensors that could be used for muscle-computer interfaces.

human physiology to apply sensors, measure activity in specific muscles, and infer intended or unintended movements from recorded data.

While this work has been quite successful, we argue that an untapped use of muscle-sensing technologies is in decoding a rich set of gestures in a relatively unobtrusive manner. Specifically, we explore the feasibility of muscle-computer interfaces (muCIs) as an interaction methodology that enables input by directly sensing and decoding human muscular activity.

Since muCIs rely on sensing relatively subtle muscle activity which sometimes leads to little or no perceivable external movement, they could enable more subtle and implicit forms of human computer interaction than those dependent on data streams traditionally sensed or used by humans to manipulate our environment or communicate with one another. Further, since we envisage that muCI sensors can essentially be worn on the body much like a watch or jewelry (see Figure 1), they provide an always-available and highly personalized input technology. This level of access to a computer input stream will likely be desirable as computing is performed in an increasingly greater variety of environments where traditional input devices may not be easily accessible.

As we expect such interfaces to be used largely by healthy individuals in real-world settings, work in this space poses several unique challenges and opportunities. First, while it is normal for medical applications to require time-consuming expert placement of sensors in order to optimally isolate muscle activity [2], end-users will not have the time, inclination, or expertise to do this. Hence, the equipment must be easy to set up and configure. Ideally, it should also be in a form factor that is unobtrusive and perhaps even fashionable to wear. Second, while work on current applications has been careful to eliminate 'noise' such as skin movement in measurements, we believe that our target environments make this extremely difficult. Fortunately, our applications instead allow us to exploit such artifacts as useful 'signal.' As we will show in this paper, this allows us to make more consistent and precise activity inferences than those previously reported (e.g., [23]).

The work presented in this paper is an initial step in exploring how EMG technology can be used as an input modality. We contribute an articulation of opportunities within this domain, and present an experiment showing that it is indeed possible to accurately classify various finger gestures across multiple fingers, including lifting and tapping, application of three levels of finger pressure, and moving to two positions using only 10 sensors placed in a band around the upper forearm. Note that at present, our intent is not to investigate classification of all possible gestures or to delve into the minutiae surrounding issues involved in building real muCI applications. While such efforts are clearly worthy of attention, it makes little sense to embark upon them without first understanding the potential and limits of EMG

sensing in this context. As such, our work is focused on providing foundational data demonstrating that using EMG for muCIs is feasible. We conclude by discussing the implication of our results to the designs of future muCIs.

BACKGROUND AND RELATED WORK

Sensing Muscles with EMG

Human skeletal muscles are made up of muscle fibers attached to bone by tendons. These muscles contract to create skeletal movement. To contract a muscle, the brain sends an electrical signal through the nervous system to motor neurons. These motor neurons then transmit electrical impulses known as *action potentials* to the adjoining muscle fibers, causing the muscle fibers to contract. The combination of a motor neuron and the attached muscle fibers are known as a *motor unit*. Each muscle is made up of many motor units. During muscle contraction, some subset of a muscle's motor units is activated. The sum of all the electrical activity in a motor unit during contraction is referred to as a *motor unit action potential* (MUAP).

Electromyography (EMG) measures the MUAP as an electrical potential between a ground electrode and a sensor electrode. EMG can measure signals either directly within the muscle (invasive EMG) or on the skin above a muscle (surface EMG). Invasive EMG is very accurate in sensing muscle activation, but is impractical for human-computer interaction applications as it requires needle electrodes to be inserted through the skin and directly into the muscle fibers. Surface EMG, while less accurate, only requires that conductive sensors be placed on the surface of the skin. Surface EMG is fundamentally noisier than invasive EMG since MUAPs must pass though body tissues such as fat and skin before they can be captured by a sensor on the surface. Due to the high sensitivity of EMG sensors required to detect these signals, they also typically detect other electrical phenomena such as activity from other muscles, skin movement over muscles, and environmental noise. For more information on the state-of-the-art in surface electromyography, see Merletti and Parker [17]. In our work, we explore the use of surface EMG for muscle sensing, and imagine people wearing future muscle-computer interaction devices as a small strap or band of sensors slid on to the upper forearm (see Figure 1).

The EMG signal is an electrical potential, or voltage, changing over time. The raw signal is an oscillating wave with an amplitude increase during muscle activation. Most of the power of this signal is contained in the frequency range of 5 to 250 Hz [17]. A typical statistic computed over the raw EMG signal for diagnosis of muscle activity is the windowed root mean squared (RMS) amplitude of the measured potential. This measure has typically been employed for diagnostic purposes such as evaluating muscle function during rehabilitation after a surgery or for measuring muscle activation to assess gait [13]. RMS amplitude is a rough metric for how active a muscle is at a given point in time. For a full review of processing techniques used in previous work, see Naik et al. [19].

EMG for Prosthetics

For over three decades, researchers have been working on using EMG as a means for amputees to use remaining muscles to control prosthetic limbs [9,11]. Most research in this domain has focused on using the muscles of the upper arms and shoulders to control the gross orientation and grasp of a low-degree-of-freedom prosthetic device for manipulating objects [9]. Each measured upper arm muscle is typically mapped directly to one degree of freedom of the prosthetic. For example, tricep contraction could be used for rotation while bicep flexion might close or open the prosthetic.

More recently, researchers have begun to look at the potential of using the forearm muscles in hand amputees to control a multi-fingered prosthetic hand. While we know of no fully functional hand prosthetic, this is clearly a promising new area of EMG research. One of the challenges for creating hand prosthetics is that there is not a trivial mapping of individual muscles to finger movements. Instead, many of the same muscles are used for several different fingers [23].

In tackling these problems, Jiang et al. [10] used wavelet transforms combined with a neural network to classify thumb, index finger, and middle finger movements from several EMG sensors placed on the upper arm and forearm. Similar research by Peleg et al. [20] has examined the possibility of differentiating among individual finger movements. They use auto regression features combined with a K-nearest-neighbor classifier to identify which of the five fingers is pressing a button. They fastened sensors to users' lower and upper forearms and immobilized their arms by attaching them to a board. Both of these techniques require sensors in multiple places on the arms (lower forearm, upper forearm, and upper arm) and require users to be in a fixed posture. Our work deviates from prosthetics research in that we deliberately aim for simpler sensor placement in a single band formation on the upper forearm, and we also try to classify a wider range of finger movements.

EMG for Human Computer Interaction

EMG sensing has been explored in HCI research for sensing emotion through facial muscle activity [16]. This work utilizes the state of human facial muscles as a mirror for the human emotional state. Since the user is not intentionally controlling facial muscle activity, this work also showcases *implicit* computer input using EMG. In fact, since EMG is sensitive enough to pick up activity that is too small to result in actual muscle movement, this work demonstrates the potential to capture emotional state via EMG even when there is no externally visible representation.

Along similar lines, Constanza et al. [5,6] have investigated subtle EMG-based interfaces. Users of their system cycle through an auditory voice mail menu on a cell phone by flexing one or both biceps. They demonstrate that people can use this input method to discretely interact with a device without being detected by others.

Wheeler, et al. [24] explored using EMG sensors on the forearm to recognize joystick movement and Naik et al.

[19] built a system that used EMG sensors distributed across the forearm to classify wrist, finger, as well as combined wrist and finger flexion. While these systems attain relatively good results, we believe that there are opportunities for extending gesture detection beyond the gross movements explored in their work. Also, we believe that wrist movements are only suitable in a subset of applications, and that finger activity will be a more natural mechanism for interacting with most computer systems. Other researchers have created EMG input devices requiring only finger movement in order to explore the space of finer motor movement [15]. However, this work requires that devices be worn directly on the fingers, which is somewhat intrusive and might interfere with normal finger activity. We seek to expand on this prior work with new techniques for higher-fidelity input via forearm muscles.

EXPERIMENT

Goals

Previous EMG-based input methods have classified gross movements such as wrist flexion or bicep activation, and hand prosthetics research has placed sensors on the forearm to detect finger movements. However, these approaches involve a restrictive setup procedure including fixing the hand to a board or placing sensors at many places on the arm in order to recognize only a few finger movements.

The high-level goal of our present work is to determine whether muCIs are even feasible using EMG technology. We seek to employ EMG technology in such a way that muscle-computer interaction can be comfortable, unobtrusive, and useful for computer input.

Our approach is to place EMG sensors in a narrow band formation on the upper forearm. We envision this eventually becoming a thin wireless band worn just below the elbow (see Figure 1). As a starting point for muCI research, we describe four sets of finger gestures and attempt to classify them computationally. We compute three simple sets of features over an eight-channel EMG signal. These features are then used for classification in an unmodified algorithm distributed in an academic machine learning toolkit [25].

We evaluate the extent to which we can classify gestures in this new muCI input space through a laboratory experiment in which participants performed these gestures. It is important to note that we deliberately chose a simple EMG setup and off-the-shelf machine learning techniques in order to determine a lower bound on classification accuracy and ensure that our techniques are accessible to other researchers. If this simple setup results in reasonable classification ability, we should only be able to do better with improved hardware and more sophisticated classification algorithms.

Tasks

Participants performed four distinct sets of finger gestures. These sets were created in order to broadly explore how well our simple setup and analysis techniques could discriminate among various characteristics in finger movements that might typically be useful in real interface applications.

| (a) Position | (c) Tap |
| (b) Pressure | (d) Lift |
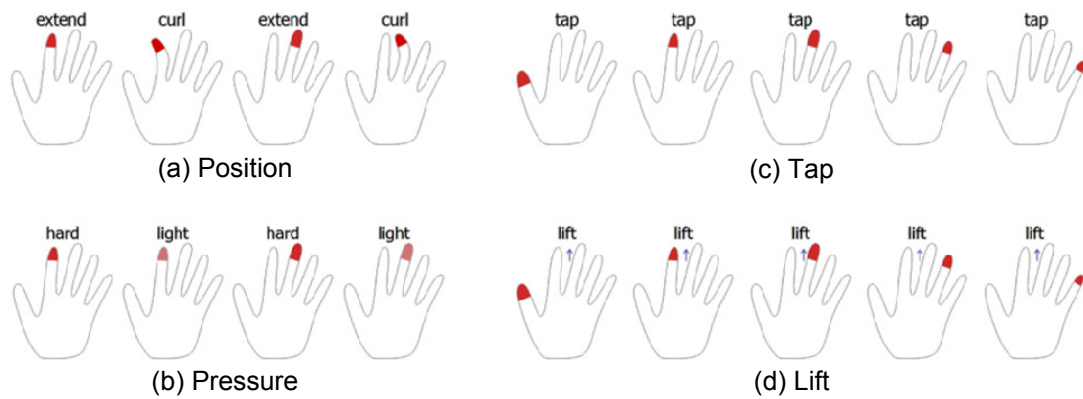
Figure 2: Stimuli representing the four sets of finger gestures tested during the experiment.

Specifically, we tested whether we could classify position and pressure of finger presses, as well as differentiate among all fingers during tapping and lifting (see Figure 2).

Gesture sets that allow the independent use of all five fingers are particularly interesting in this exploration. Physiologically, since not all the fingers have independent muscle bundles that run down through the forearm (e.g., control of the thumb is mostly done through a strand that runs only to the palm and wrist), we expected this to be a very difficult classification task, as demonstrated by multiple researchers who have tried before us and resorted to using only gross movements for control. However, we believe that the ability to classify among gestures performed by all five fingers independently is powerful in designing muCIs. Of course, not all gestures are easily performed independently by all fingers, and our choice of which fingers to study for particular gestures reflects these differences among fingers.

Each participant began each of the tasks with his hand in a relaxed palm-down position on the table, which we refer to as the *rest* position. For most participants, this meant that fingers were mostly extended, but slightly curled. We asked participants to return to this rest position between gestures.

Position: Index & Middle Finger
The four gestures in the *Position* set involved touching a surface with either the index or middle finger in one of two positions: extended or curled (Figure 2a). Touching in the extended position is characterized as touching the surface with a finger while the hand is in the rest position. A curled position involves sliding the pad of the finger closer to the palm, which also causes the knuckle to rise, and touching the surface in that position. Participants were instructed to touch the surface firmly but not excessively hard.

Pressure: Index & Middle Finger
The four gestures in the *Pressure* set involved pressing lightly or firmly on a surface with either the index or middle finger (Figure 2b). This was done in the extended position (i.e., from rest). We told participants that when "pressing lightly", they should apply enough pressure to dent a tomato, while "pressing hard" would break the skin of the tomato. Including the *rest* state, this gesture set has three levels of pressure.

Tap: Each of the Five Fingers
The five gestures in the *Tap* set involved individually tapping each of the five fingers on a surface from the rest position (Figure 2c). This involved raising the finger slightly and then returning it onto the surface. Participants were told to do this comfortably and not to exaggerate the gesture.

Lift: Each of the Five Fingers
The five gestures in the *Lift* set involved individually raising each of the five fingers in the air and holding it there (Figure 2d). Participants were told to raise the finger only to the point where further movement would require additional force, and not to exert any more force than was required to life the finger off the table.

Equipment and Setup
We used a *BioSemi Active Two* system for performing EMG sensing on participants' forearms [3]. This device employs a two-electrode active grounding system that drives the average potential of the participant as close as possible to the amplifier ground and reduces noise levels. The device samples eight sensor channels (labeled EX1 through EX8 in Figure 3) at 2048 Hz. While the capability of the Active Two to sense EMG is secondary to its primary electroencephalograph (EEG) function, the signals from this research device are comparable to most mid-end commercial surface EMG units used in medical settings. Sensor data was recorded to hard disk via a USB connection.

Before placing sensors on participants, we had them clean their upper forearms with an abrasive skin scrub, which helped ensure good conductivity and easy attachment. We
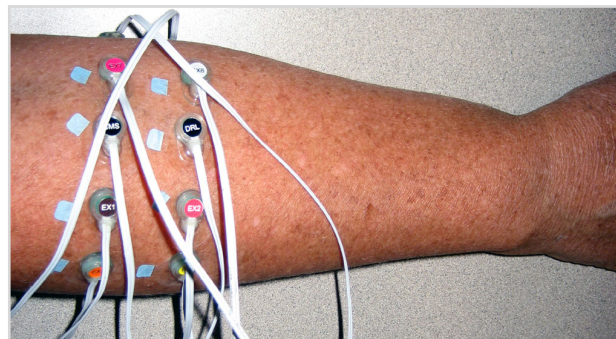
Figure 3: Sensors placed in a narrow band around a participant's arm. Picture taken from actual experiment.

then applied conductive gel to each sensor and attached them to the participants' skin with a small adhesive circle around the gel. This cleaning procedure and the use of conductive gel are simply artifacts of the particular EMG equipment we used. Both can be obviated if dry electrodes, available with other manufacturers' equipment, are used instead (e.g., see www.neumed.com). Dry electrodes would clearly be more appropriate for actual muCI applications.

Traditionally, EMG sensors are placed on the muscle belly, the largest mass of muscle that powers the particular movement of interest, in order to robustly measure the action potentials. Pairs of sensors are typically placed about an inch (~2.5cm) or so apart, in line with the muscle fibers. Most current work also alludes to requiring intimate knowledge of surface muscle anatomy in order to get good sensor placement and clean signal.

In our pilot studies, we found that we could get reasonable signal even when we only approximately placed the sensors, especially when we had multiple sets of them placed around the forearm. Hence we decided to place pairs of the eight sensors and the two ground electrodes in a narrow ring around each participant's upper right forearm (see Figure 3). We evenly spaced each pair around the portion of the forearm that did not rest on the table, with the ground electrodes in the middle. This configuration has encouraging implications for potential form factors of an approximately-placed armband EMG device (as illustrated in Figure 1). We draped sensor cables over the tops of participants' arms, allowing them to rest their arms comfortably on the table. Again, the presence of these cables was an artifact of our device, and wireless units are already commercially available [12]. On average, setup took a little less than 10 minutes.

Participants found a comfortable sitting position and viewed visual stimuli on a 21″ Samsung SyncMaster 214B LCD display. While we advised participants to minimize extraneous movements, we did not interrupt or stop them when they moved. Many participants occasionally stretched their hands, rocked their chairs back and forth, or tapped their feet. We believe that testing in such an environment is important as it mimics the realistic usage conditions of future muscle-computer interfaces.

Design and Procedure
A within-participant design was used, with each participant performing all four tasks: Position, Pressure, Tap, and Lift. We counter-balanced the order of tasks across participants using a Latin square design.

A randomized block design was used within each task. Each block consisted of a single repetition of each gesture within a set, presented in random order. We refer to each instance of a performed gesture as a trial. Participants performed multiple consecutive blocks for each task before proceeding to the next task. This block design is important as it ensures that gestures were relatively well distributed in

time and that temporal bias (e.g., measurement drift) does not artificially inflate classification results.

For each trial, the participant was presented a visual stimulus that indicated the gesture to be performed (still images from Figure 2 presented on a computer screen). Participants were told to perform the appropriate gesture for the duration the stimulus stayed on the screen. In the Position, Pressure, and Lift tasks, stimuli were presented for two seconds. Tap stimuli were only shown for three quarters of a second, because a tap is a discrete gesture and cannot be "held down" or "held up." Participants were told to only tap their finger once per stimulus. A random delay between one and two seconds was inserted between stimuli to allow participants to return their hand to the rest position and refocus attention. Before each task, participants were given instructions and had approximately two minutes to practice performing the gestures with the prompting of the stimuli.

For each of the Pressure, Lift, and Tap tasks, participants completed 50 blocks with a two-minute break after 25 blocks were completed. Thus, each gesture for each of those tasks was performed 50 times by each user. In our pilot studies, we found the Position gesture set to be the most difficult to classify. Hence, we collected 75 blocks instead, resulting in 75 trials for each Position gesture, in order to have more training data for classification.

Participants
Thirteen individuals (8 female) from the broader community volunteered for the experiment. Participants ranged from 20 to 63 years of age with an average age of 46. Participants were between 5′1″ and 6′1″ (1.55m to 1.85m) and weighed from 125 lbs to over 225 lbs (57kg to over 102kg). Most were daily computer users and played video games infrequently. None of the participants reported any existing muscular conditions or skin allergies, and all were right-handed. None were colorblind and all had 20/20 or corrected-to-20/20 vision. The experiment took approximately 90 minutes and participants received a software gratuity.

One participant remarked several times during and after the experiment that he "zoned out" and did not respond to the stimulus for sizable stretches of time. As such, we did not use this participant's data, discarding it before we began our analysis. In the following sections, we report on analysis of the data collected from the remaining twelve participants.

Data Analysis Techniques
In order to classify the signals measured with our EMG setup, we first perform some basic signal processing to transform the time series data into a time-independent data set. We then compute a set of features, which we use to train a support vector machine (SVM) [4] and perform the classification. We discuss two metrics that quantify classification quality, one based on the accuracy with individual samples, and the other based on classifying entire trials. These steps are described in the following subsections.

Basic Signal Processing

We first convert the raw EMG data into a time-independent dataset, which is generally better-suited to most machine learning algorithms. To do this, we adopt a technique used in EEG work [11]: we divide the raw EMG signal into 250ms segments and treat each 250ms segment as a single *sample* of EMG data. We then apply a band-pass filter to each sample between 2 Hz and 102 Hz, as pilot studies indicated that this is where most of the useful signal resides. We also filter out the 55 Hz to 65 Hz band in order to remove the 60 Hz noise that exists in most computing environments (e.g., from power lines and appliances).

Feature Generation

For each 250ms sample, we generated three classes of features, which we use for training and testing the classifier. These classes were chosen based on prior work suggesting that they may be discriminative of various activities within electroencephalography and EMG signals [14,17,22].

- *Root Mean Square* (RMS) amplitude of the EMG potential is indicative of the amplitude of muscle activity close to a particular sensor. We first fully rectify the signal by taking its absolute value and then compute the RMS amplitude of each of the 8 channels. From these 8 base RMS features, we create another 28 by taking the ratios of the base RMS values between each pair of the channels. These ratios make the feature space more expressive by representing relationships between channels, rather than treating each as being completely independent.

- *Frequency Energy* is indicative of the firing rate of muscle activity. The energy is often thought to be significantly affected by muscle energy and fatigue [17], and our pilot studies suggest it has high discriminative power in the activity space as well. To derive the frequency energy feature, we compute the fast Fourier transform (FFT) for each sample and square the FFT amplitude, which gives the energy at each frequency, and sum the energy of all channels into 10 Hz bins. This yields 10 frequency energy features for each sample.

- *Phase Coherence* measures the extent to which the EMG channels in a sample have a fixed relationship to each other (e.g., firing in similar manner). This is used extensively in electroencephalography work [14] and pilot studies suggest it has discriminative power for EMG as well. As with the RMS values, we create 28 features by taking the ratios of the average phase between all channel pairs.

These calculations result in 74 features per sample in total.

Classification

In our pilot studies, we explored several machine learning techniques for classifying the EMG signal into gestures. In those studies, we found that SVMs seemed to perform well [4]. SVMs are a set of supervised machine learning methods, which take a set of labeled training data and create a function that can then be used to predict the labels of unlabelled data. The labeled training data typically consist of input vectors of feature values and desired outputs. At a high level, SVMs map the input vectors to a high-dimensional space and attempt to create a set of maximal separating hyperplanes between the output variables, or classes. For our experiment, we used the Sequential Minimal Optimization (SMO) version of support vector machines [21] as implemented in the Weka toolkit [25]. SMO is a fast method for training SVMs that breaks the typically large quadratic programming problem into a series of the smallest possible problems which are then solved analytically. This optimization minimizes computation time and memory use.

We used the default parameters for Weka's implementation of the SMO algorithm. Specifically, we set Weka's complexity parameter to 1.0 and the epsilon for round-off error at 1.0E-12, and we used a polynomial kernel function with exponent 1 and tolerance parameter of 0.001.

In labeling the training data we collected, we could have used values derived directly from the stimulus presentation (i.e., when the stimulus was showing a particular value, the data would be labeled accordingly), but we observed that participants responded to the stimulus with varying delay. Therefore, labeling based on the presented stimuli is more accurate toward the end of the stimulus presentation period. Hence we discard the first four 250ms samples (a total of one second per trial) and use only the next four samples for training. To acquire rest samples, we take the single sample that immediately preceded each gesture stimulus, as this was projected to be the time in which the hand had maximally recovered from the previous gesture but had not yet begun the next. Doing this also allowed us to have equivalent amounts of data from the rest and active conditions. See Figure 4 for an illustration of the data samples used. Throughout this paper, our classifiers were trained and tested independently on data from each user from a single user session (except where stated otherwise).

Classification Results

We conducted a ten-fold cross-validation for tabulating classification accuracies for all four of our gesture sets. In each of the ten folds, 90% of the collected data was used for training and 10% was used for testing. The testing data were always taken as a single continuous chunk in time, as this is the most realistic scenario for testing such applications. Performing a random holdout set, as is traditionally
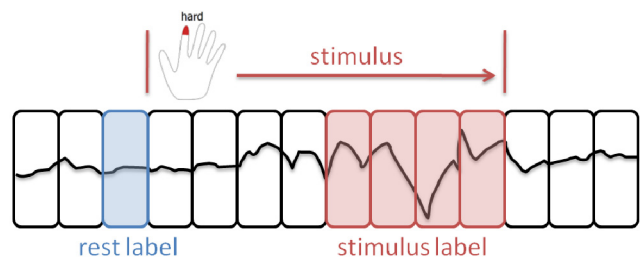
Figure 4: To mitigate variability introduced by participant reaction time, we trained and tested using only the sample before each stimulus was shown and the last four samples during stimulus presentation.
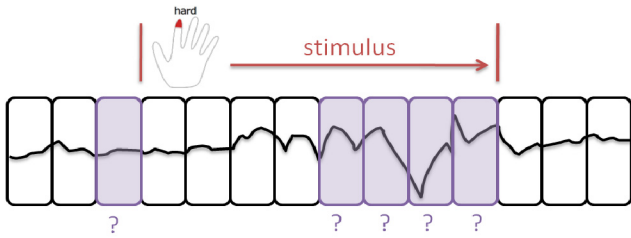
Figure 5: In single-sample classification, we classify each of the data samples independently.

Figure 6: In whole-trial classification, we take a majority vote among classifications of individual samples in the largest non-rest region.

done, could have artificially boosted our results since test data samples would have been temporally adjacent to training samples. Using the training and testing methodology described above, we evaluated our classification accuracies for the four gesture sets.

Single-Sample Classification
Our first metric is the accuracy of single-sample classification. This metric provides a sense of the overall performance of the classifier. For the same reasons described above with respect to our training data, we only classify the sample immediately before each stimulus and the four samples at the end of the stimulus (see Figure 5). We do not try to classify the first four samples within the stimulus presentation period because we do not know when the physical gesture actually began due to variance in participant response times. The four samples we do classify are treated independently of each other. Classifying the sample immediately before stimulus presentation allows us to include the non-active or rest state as a condition in our classification.

In classifying single samples, our Position classifier performed at an average accuracy of 71% (sd: 9.0%) while Pressure classified at about 76% (sd: 6.1%). In both cases, these results are much higher than the prior probability (the expected performance of a random classifier). For both of these tasks, prior probability is 20% because the classifiers were deciding among five conditions (four gestures and rest), each with an equal number of samples.

The Tap gestures were classified at an accuracy of 75% (sd: 6.9%) and Lift at 87% (sd: 7.0%). Again, these results are significantly above the prior probability, which was 17% since they were differentiating among six conditions (five gestures and rest). Per-participant accuracies for these results are shown in Figure 7.
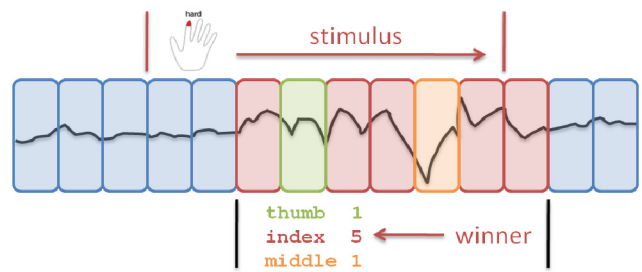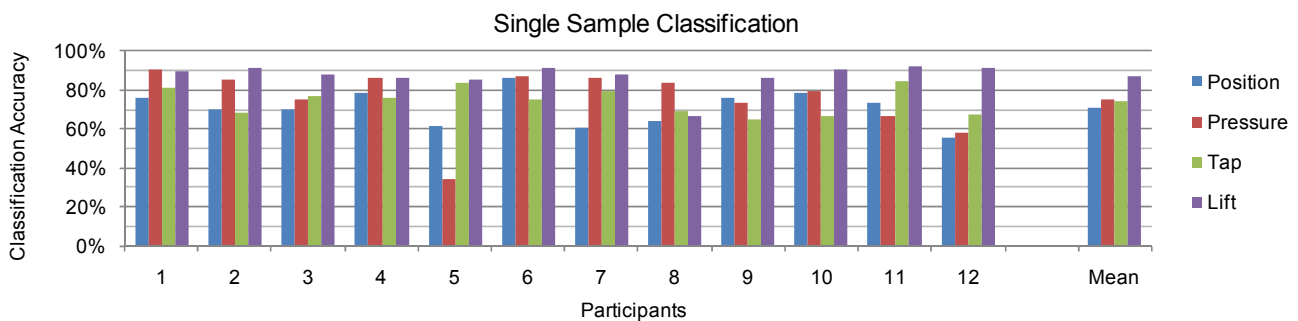
Whole-Trial Classification
Our second metric is the accuracy of whole-trial classification. In a muCI, we probably care more about classifying an entire gesture, or trial, rather than single data samples. As such, this metric provides us with a more ecologically valid measure of classifier performance. We consider a trial as the period starting three samples before stimulus presentation and ending three samples after (Figure 6). In order to classify a trial, we take a majority vote among classifications of individual samples in the largest non-rest region. When there is a tie, the vote that occurred latest wins.

Accuracy increases for all tasks when we use this more ecologically valid whole trial classification metric. With this metric, Position gestures are recognized at 78% (sd: 9.0%), Pressure at 84% (sd: 4.7%), Tap at 78% (sd: 10.5%), and Lift at 95% (sd: 6.4%). Figure 8 summarizes this data.

We note that if the classification problem is simplified by removing some of the gestures that are difficult to disambiguate, accuracy increases. For example, when all of the light pressure trials are removed from the Pressure task data, and the system only has to differentiate the firm index finger press, firm middle finger press, and rest conditions, it performs at an accuracy of over 90% (sd: 3.5%).

Quantity of Training Data and Classification Performance
In the above evaluations, we used 90% of the data collected for training. We believe that it is important to understand the time-performance tradeoffs that exist in collecting training data, which has implications for the time required to configure muCI's in the future. Hence, we analyzed our data with various amounts of training data, incrementally removing samples from the end of the training data.

Figure 7: Single-sample classification accuracies for all four sets of gestures, broken down by participant, with the mean result on the right side of the graph.
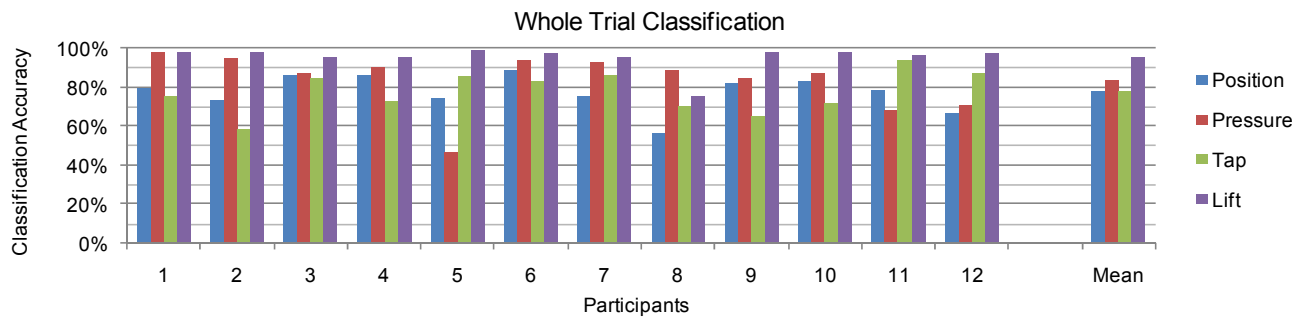
Figure 8: Whole-trial classification accuracies for all four sets of gestures, broken down by participant, with the mean result on the right side of the graph.

We found, not surprisingly, that decreasing the amount of training data decreases the accuracy of the classifier (Figure 9). However, we were pleasantly surprised at how slowly the results degrade as training data is reduced and how robust they are to the amount of training data available. In fact, each classifier's accuracy decreases only about 20% as we reduce the amount of training from 45 blocks to just 7 blocks. For example, Lift gestures are still classified at almost 70% accuracy with only two minutes of training data.

Discussion

Noisy Labels
One of the most significant limitations on our current classification results was the degree of inaccuracy in our labels when we trained and tested the gesture classifier. There were two main sources of noise in the labels.

The first type of noise is introduced by a variance in delay between stimulus presentation and participant response. This delay created many mislabeled data samples since the stimulus presentation could not be assumed to match the actual muscular response. While we mitigated the effect of this noise by using only the latter samples within a trial, we do not think this resulted in entirely correct labels, as variance between participants was quite large, with at least one participant who seemed to respond well beyond the one-second window of data that we did not classify.

The second source of noisy labels comes from participant error: in some cases, participants performed a different gesture than that instructed by the stimulus. Some participants remarked that they would try to anticipate upcoming stimuli and would make mistakes when they did not guess correctly. Hence, some percentage of our classification "errors" may actually be correct based on the gesture actually performed, but we do not know how frequently this occurred.

Both these problems could have been alleviated had we instrumented the participant or surface with physical switches or gesture sensors so we would know exactly when a finger pressed down or lifted. We choose not to implement such a setup because we wanted participants to perform gestures unencumbered and in a manner that was comfortable and intuitive for them, to maintain ecological validity with respect to our proposed application scenarios. While performing a study with such a control remains future work, we are encouraged by the surprisingly good results even in the presence of what we believe are fairly noisy labels, which only hurt our current results. We also believe that many of these problems would be minimized in an interactive system where users would have feedback indicating the system's classifications of their actions.

What Are We Measuring?
While traditional EMG work has been careful to ensure that the measured signal is derived from muscle unit action potentials, it is important to recall that these sensors do also detect other electrical signals. These signals include firing from distant muscles, such as the heart, environmental noise, and most importantly gross body movements. Many of these body movements cause the skin on which the sensors are attached to move as well, creating electrical signal.
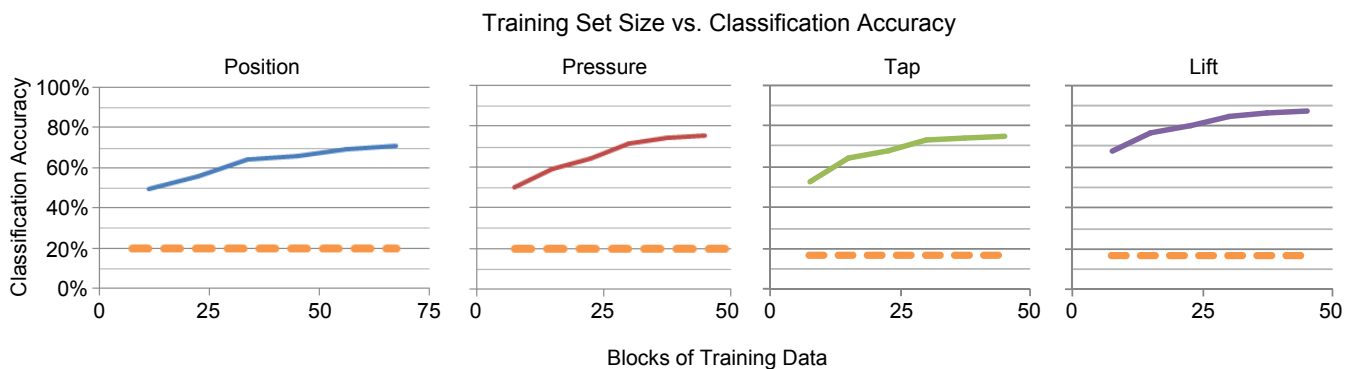
Figure 9: Decreasing the amount of training data degrades classification accuracy for all four gesture sets, but performance remains surprisingly high even with very little training data. The dotted orange line represents chance.

While many of these extraneous signals would be considered too unreliable for use in traditional medical settings, we believe that some portion of our measured signals may actually be due to some of these "artifacts", but assert that these signals are perfectly legitimate, if not desirable, for use in muCIs that aim to classify overall activity. We believe that this approach of utilizing every bit of available signal is one of the reasons we are able to classify independent movement of all five fingers with such high accuracy, even though not all the fingers have muscle bundles that run down the forearm.

Cross-User Classification to Reduce Training Time
In the results presented above, classifiers were trained and tested independently on data from each user with a fixed-placement of electrodes. That is, a new classification model was created for each user. In this scenario, every user has to train the system before it can recognize his or her gestures with each placement of electrodes.

To extend this result, we explored how well we could create cross-user models: models that are trained on data from one set of users and applied to classification on another. Being able to do this has implications for the robustness and invariance of the features we are using, and also for the potential of creating systems that require little or no user-specific training data. As an initial exploration of this problem space, we performed a twelve-fold cross validation in which we held out data from one user and trained on data from the remaining eleven. We attained reasonable though not stellar results. For example, we classify Lift gestures at an average of 57% accuracy (chance was 17%). It is not surprising that this accuracy is lower than for single-user models, especially since we did not do anything in sensor placement, experimental design, or data analysis to provide specifically for cross-user transfer. Despite these limitations, the classifier performed considerably better than chance, suggesting that there exists potential for building cross-user classifiers in this domain.

A more sophisticated approach to cross-user modeling might involve cross-validating internally within the training data to select the most invariant features from which to build the classifier. Exploring this remains future work.

IMPLICATIONS FOR THE DESIGN OF muCIs
The classification results presented in this paper are based on single-handed input. A similar sensor band could be placed on the other forearm to provide even more input possibilities. Since the classifiers would work nearly independently, we expect that this would approximately double the gesture space with little or no drop in accuracy. In fact, if we consider compound gestures involving both hands, we could imagine the input space growing quite significantly.

We believe the form factor of a wireless band just below the elbow is feasible in the near future for muscle-computer interfaces. The components for such a system are relatively inexpensive and wireless EMG products already exist (e.g., [11]). A band could be unobtrusive and worn either openly as a fashion accessory, or concealed beneath clothing. In either scenario, it would provide an input capability where traditional input devices are not feasible or desirable. We note that this form factor does not guarantee identical sensor placement as the user removes and replaces the armband; the results presented in this paper assume that sensors did not move between training and decoding. Further work will be necessary to establish the training paradigms or mechanical constraints that will be necessary to generalize to a removable form factor.

The index and middle finger are arguably the most viable initial candidates for use in muCI applications that rely on position-based and pressure-based gestures, as it is difficult to independently control the fourth and fifth fingers for such gestures, and the thumb's abilities in this regard are qualitatively quite different from the other fingers. As such, our work focused on the index and middle fingers when evaluating position and pressure. The promisingly high classification accuracies we obtained indicate that simple muCIs could be designed that allow for the control of one or two input channels using the index and middle fingers. Examples include two-position sliders or switches controlled by extending and curling these two fingers, and three-state buttons controlled by varying the pressure exerted by these fingers (with button states corresponding to the "rest", "pressing lightly" and "pressing hard" gestures we evaluated in our experiment). The beauty of doing this with a muCI is that these gestures can be performed *anywhere* and against any surface, whether a table one happens to be sitting at, one's own lap, or a car's steering wheel.

In contrast to position and pressure, simpler gestures such as tapping and lifting are reasonably independently performed by all fingers, hence our experiment evaluated all five fingers for these two gestures. Tapping and lifting with all five fingers are of particular interest, as our results indicate that we could conceivably design a muCI that allowed any surface to be turned into a virtual keyboard operable by all five fingers (and potentially all ten fingers if a similar sensor band was worn on both arms). This has significant implications for text input where traditional keyboards may not be viable: imagine sitting in a meeting, typing away with your fingers drumming on your lap with visual feedback obtained through your eye-glass display, while others in the room are entirely unaware of your secondary activity. In fact, visual feedback might well be provided by a standard laptop screen: while typing on a laptop's keyboard might be distracting to other meeting attendees, typing via your muCI and receiving feedback via a standard laptop display would likely not.

CONCLUSIONS AND FUTURE WORK
Our techniques and experiment show potential for unobtrusively sensing and decoding muscular activity for computer input. It is important to note that our results, with classification accuracies as high as 95% in some conditions, were achieved with off-the-shelf machine learning techniques, and with casually placed EMG sensors. It is reasonable to

expect that even better classification accuracies could be achieved with more highly tuned machine learning algorithms and purpose-built sensors. We note, however that our results were obtained with participants in a relatively sedentary position and gestures were performed beginning from a well-defined rest state. It remains to be investigated how classification accuracy degrades in situations where the user is less sedentary or where gestures are performed in quick succession with possible overlap.

Our results provide a baseline that we hope will spur further research geared towards realizing a fully deployable muCI in the near future. Apart from refining the hardware technology and classification techniques, an interesting next step would be to investigate interaction scenarios enabled by this style of sensing. Of particular interest is how users would perform with muCIs that can be used unobtrusively in any locale, but that lack the passive haptic feedback of most current physical input technologies. We believe that with the right design of interaction techniques coupled with appropriate visual and auditory feedback, the impact of not having haptic feedback can be successfully mitigated, but this hypothesis remains to be formally investigated.

ACKNOWLEDGEMENTS

REFERENCES

1. Barry, D.T., Gordon, K.E. & Hinton, G.G. (1990). Acoustic and surface EMG diagnosis of pediatric muscle disease. *Muscle & Nerve, 13(4),* 286-290.

2. Basmajian, J.V., & De Luca, C.J. (1985). *Muscles alive: Their functions revealed by electromyography,* 5th edition. Williams & Wilkins: Fort Worth, Texas.

3. Biosemi. http://www.biosemi.com.

4. Burges, C. (1998). A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery, 2,* 121-167.

5. Costanza, E., Inverso, S.A., & Allen, R. (2005). Toward subtle intimate interfaces for mobile devices using an EMG controller. *ACM CHI Conference,* 481-489.

6. Costanza, E., Inverso, S.A., Allen, R., & Maes, P. (2007). Intimate interfaces in action: Assessing the usability and subtlety of EMG-based motionless gestures. *ACM CHI Conference,* 819-828.

7. Hartmann, B., Benson, M., Junger, A., Quinzio, L., Röhrig, R., Fengler B., Färber, U.W., Wille B., & Hempelmann, G. (2004). Computer keyboard and mouse as a reservoir of pathogens in intensive care unit. *Journal of Clinical Monitoring and Computing, 18(1),* 7-12.

8. Hernandez, A., Katoh, R., Yokoi, H., & Yu, W. (2006). Development of a multi-DOF electromyography prosthetic system using the adaptive joint mechanism. *Applied Bionics and Biomechanics, 3(2),* 101-112.

9. Jacobsen, S.C. & Jerard, R.B. (1974). Computational requirements for control of the Utah arm. *ACM Annual Conference,* 149-155.

10. Jiang, M.W., Wang, R.C., Wang, J.Z., & Jin, D.W. (2005). A method of recognizing finger motion using wavelet transform of surface EMG signal. *Intl Conf. of the IEEE EMBS, 1(7),* 2672-2674.

11. Kiguchi, K., Tanaka, T., & Fukuda, T. (2004). Neuro-fuzzy control of a robotic exoskeleton with EMG signals. *IEEE Trans. on Fuzzy Systems, 12(4),* 481-490.

12. Kine Biomedical Engineering. http://www.kine.is.

13. Lanyi, X., & Adler, A. (2004). An improved method for muscle activation detection during gait. *Canadian Conf. on Electrical and Computer Engineering,* 357-360.

14. Lee, J.C., & Tan, D.S. (2006). Using a low-cost encephalograph for task classification in HCI research. *ACM UIST Symposium,* 81-90.

15. Manabe, H., Hiraiwa, A. & Sugimura, T. (2003). A ring-shaped EMG measurement system for applying to user interface. *Intl. Conf. of the IEEE EMBS,* 3020-3023.

16. Mandryk, R.L., Atkins, M.S., & Inkpen, K.M. (2006). A continuous and objective evaluation of emotional experience with interactive play environments. *ACM CHI Conference,* 1027-1036.

17. Merletti, R., & Parker, P.A. (2004). *Electromyography: Physiology, engineering, and noninvasive applications.* John Wiley & Sons: Hoboken, New Jersey.

18. Moseley, J.B., Jobe, F.W., Pink, M., Perry, J., & Tibone, J. (1992). EMG analysis of the scapular muscles during a shoulder rehabilitation program. *The American Journal of Sports Medicine, 20,* 128-134.

19. Naik, G.R., Kumar, D.K., Singh, V.P., & Palaniswami, M. (2006). Hand gestures for HCI using ICA of EMG. *HCSNet Workshop on the Use of Vision in HCI,* 67-72.

20. Peleg, D., Braiman, E., Yom-Tov, E., & Inbar, G. (2002). Classification of finger activation for use in a robotic prosthesis arm. *IEEE Trans. on Neural Systems and Rehabilitation Engineering, 10(4),* 290-293.

21. Platt, J. (1998). Sequential Minimal Optimization: A fast algorithm for training support vector machines. *Microsoft Research Technical Report MSR-TR-98-14.*

22. Raez, M.B.I., Hussain, M.S., & Mohd-Yasin, F. (2006). Techniques of EMG signal analysis: detection, processing, classification, and applications. *Biological Procedures Online, 8,* 11-35.

23. Schieber, M.H. (1995). Muscular production of individual finger movements: the roles of extrinsic finger muscles. *Journal of Neuroscience, 15(1),* 248-297.

24. Wheeler, K.R. & Jorgensen, C.C. Gestures as input: Neuroelectric joysticks and keyboards. *IEEE Pervasive Computing, 2(2),* 56-61.

25. Witten, I.H., & Frank, E. (2005). *Data mining: Practical machine learning tools and techniques,* 2nd Ed., Morgan Kaufmann: San Francisco.