Exploring the possibility of establishing baseline EMG signals for simple hand gestures

Prepared by Benjamin Olsen

Abstract - **This paper explores a method for checking a population of EMG signals for establishing a benchmark baseline signal for hand grip gestures (spherical, tip, palmar, lateral, cylindrical, and hook) using only two EMG electrodes placed on the Flexor Carpi Ulnaris and Extensor Carpi Radialis, Longus and Brevis, and one reference electrode. The raw signals are decomposed and the mean amplitude over adjacent sample windows of 50 is extracted. A baseline for each of all samples grouped by hand grip gesture is computed by averaging the signal across each grip and computing a 95% confidence interval for the moving mean. To test the hypothesis that the population observed could be considered to be a representative sample of a greater population, the statistical difference of the signals grouped by grip and the statistical difference of the signals within each grip grouped by subject is tested utilizing a Kruskal–Wallis one-way ANOVA test, which in both cases indicates that the medians of each group are statistically different ($p<0.01$). This indicates that while the signals by hand grip are statistically different, the signals within that grip population (grouped by subject) are actually part of different populations, causing us to reject the hypothesis that the data is representative of a larger population and can be reliably used to compute a reference baseline.**

I. INTRODUCTION

Documentation is an important pillar of healthcare because it drives medical advancement, encourages growth as a community through accountability, and provides insurance companies with enough proof for them to reimburse medical treatments. In what may have been a previously unintended effect of healthcare insurance is that the administration of certain medical treatment is indirectly driven by what health insurance companies deem to be "reimbursable." While whether or not this relationship between healthcare providers and insurance companies is ultimately beneficial to healthcare as a whole is open to debate, the reality of the situation is that no procedure can become widely used if its outcomes cannot be acceptably documented. This is an issue that has withheld entire professions from breaking into the healthcare industry because insurance will not cover treatments if their outcomes cannot be convincingly proved. In terms of measuring skeletal muscular actions, usage of EMG data has thus far been popularly used as a diagnostic tool, but has not been considered for documentation purposes mostly due to the noisiness of the data in EMGs. However, new advances in the area of signal decomposition may make it possible to establish a population baseline for EMG data associated with certain musculoskeletal motions, opening the possibility of using EMGs to document therapeutic progression of muscle rehabilitation. This would make documentation much easier to collect, and potentially make the healthcare industry more accessible to new therapies. This project is an attempt to demonstrate the possibility for using a small population of easily collected signals to establish a baseline for which other signals can be compared to.

II. BACKGROUND DETAILS

Electromyography (EMG), according to Wikipedia, is an electrodiagnostic medicine technique for evaluating and recording the electrical activity produced by skeletal muscles by using electrodes to detect the electric potential generated by muscle cells when these cells are electrically or neurologically activated. The signals can be analyzed to detect medical abnormalities, activation level, or recruitment order, or to analyze the biomechanics of human movement. For the purposes of this project, surface EMG signals were collected from two sites (the Flexor Carpi Ulnaris and Extensor Carpi Radialis, Longus and Brevis group) with the

reference electrode place between the two sites to create a two channel EMG signal.

Activity recorded in EMG signals (measured in microvolt changes between an electrode and reference electrode known as "activation units", or a.u.) exhibits a linear relationship to muscle contraction and number of contracted muscles, meaning that the stronger the muscle contraction and the higher the number of muscles activated, the higher the amplitude of the signal recording will jump.

III. MATERIALS

The data in question comes from: https://archive.ics.uci.edu/ml/datasets/sEMG+for+Basic+Hand+movements

A. Instrumentation:

The data were collected at a sampling rate of 500 Hz, using as a programming kernel the National Instruments (NI) Labview. The signals were band-pass filtered using a Butterworth Band Pass filter with low and high cutoff at 15 Hz and 500 Hz respectively and a notch filter at 50Hz to eliminate line interference artifacts. The hardware that was used was an NI analog/digital conversion card NI USB- 009, mounted on a PC. The signal was taken from two Differential EMG Sensors and the signals were transmitted to a 2-channel EMG system by Delsys Bagnolia Handheld EMG Systems.

B. Protocol:

The experiments consisted of freely and repeatedly grasping of different items, which were essential to conduct the hand movements. The speed and force were intentionally left to the subject's will. There were two forearm surface EMG electrodes Flexor Carpi Ulnaris and Extensor Carpi Radialis, Longus and Brevis) held in place by elastic bands and the reference electrode in the middle, in order to gather information about the muscle activation. The subjects were asked to perform repeatedly the following six movements, which can be considered as daily hand grasps (*Figure 1)*: a) Spherical (for holding spherical tools), b) Tip (for holding small tools), c) Palmar (for grasping with palm facing the object), d) Lateral (for holding thin, flat objects), e) Cylindrical (for holding cylindrical tools), f) Hook (for supporting a heavy load). Two different databases are included: a) 5 healthy subjects (two males and three females) of the same age approximately (20 to 22-year-old) conducted the six grasps for 30 times each (the measured time is 6 sec) and b) 1 healthy subject (male, 22-year-old) conducted the six grasps for 100 times each for 3 consecutive days (the measured time is 5 sec.) (the final subject was excluded in this analysis to avoid imbalancing representation).

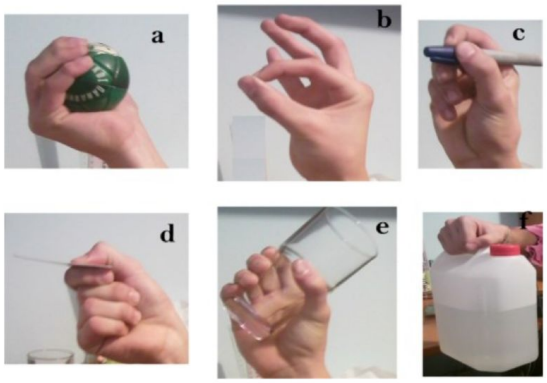

Figure 1: Depiction of each hand grip gesture.

IV. EXPERIMENTAL DESIGN

A. Data Wrangling Steps

1) Concatenate Files

Each file is a ".mat" file, one for each subject from whom data was collected. Each file contains a dictionary so that the data is organized: subject -> grip motion -> measurements. Each measurement entry is 3000 samples, and there is a total of 30 entries per grip. To follow best practices, each signal must be encoded as a separate observation, along with categorical variables to identify subject, grip, channel number, grip_channel, and sex of each observation. For lack of a better way to format the data, each array of 3000 readings will be oriented as rows and each reading over time will be a column.

2) Correcting the Centrality of the Signals

Each signal should be oriented around the x-axis, but due to error EMG signals are sometimes collected off-axis. By subtracting the mean of an array from that array, the signals can be recentered around the x-axis.

3) Rectifying the Signal

Signals measure active changes using x-axis = 0 as a reference point. Simplifying the signal to just observe the magnitude of the change involves taking the absolute value of each measurement.

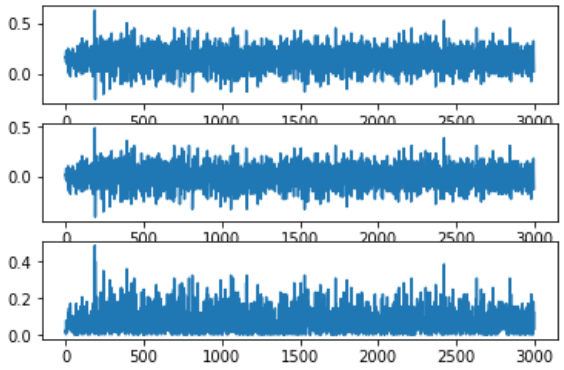

Figure 2: Raw EMG signal (top), centralized signal (middle) and rectified signal (bottom).

B. Feature Extraction

1) Calculating the Signal Envelope

A common EMG decomposition technique is to use a low pass filter to filter out extremes and capture the approximate activation of the signal over time. This is done by applying a 4th order butterworth filter with a low pass cutoff of 10 Hz (the butterworth filter requires that the nyquist frequency be calculated which is equal to the lowpass frequency divided by the sampling frequency).

2) Extracting Windowed Amplitudes

A simple way to reduce number of features while still capturing the shape of the signal is the extract the mean amplitude value over a adjacent windows of the data. In this situation the mean value has been selected to represent the activation, although the median value may be considered. By extracting the mean signal value from every 50 signals, the number of variables has been reduced from 3000 to 60 while still maintaining the signals overall shape.

3) Extracting Overlapping Windowed Amplitudes

Like before, but this time with overlapping windows instead of adjacent windows. This results in a much smoother line. Using the average of the windows gives a result even smoother than the low pass filter with reduction in necessary samples by 50x. For consistency, the windows contain a total number of 345 samples with an overlap of 300 so as to maintain 60 samples after reduction.

C. Exploratory Data Analysis

The purpose of the EDA was to consider which features would be appropriate for calculating and testing a potential baseline, and to consider which features might be appropriate for comparing new signals to those baselines in the future.

It is apparent that the enveloped signal is the closest and most digestible decomposition of the signal, capturing a lot of the variation while smoothing out the signal. The mean amplitude extracted over adjacent windows reduces the number of samples substantially while still capturing the overall shape of the signal. Finally, the mean amplitude over overlapping windows rendered the smoothest and most simplified version of the signal (*Figure 3*).

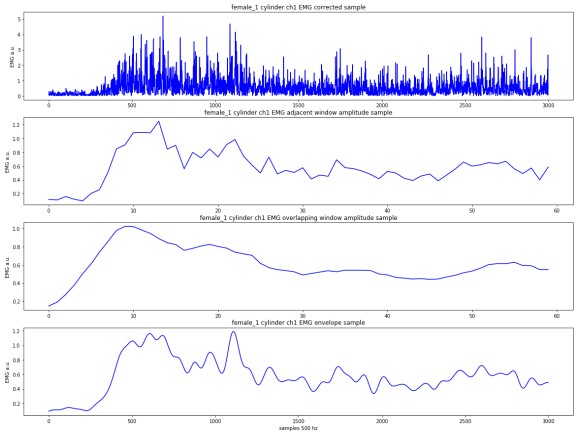

Figure 3: Raw EMG signal (top), mean amplitude over adjacent widows (2nd from top), mean amplitude over overlapping windows (3rd from top), and signal envelope (bottom).

D. Inferential Statistics

The distribution of the signals grouped by grip per channel appear to be skewed to the right. This makes a case for the as median a better measure of centrality. In this case capturing variety is a priority, so the mean is still a valuable measure of centrality in this data set, especially since in most cases the mean wasn't severely affected by outliers (*Figure 4*).

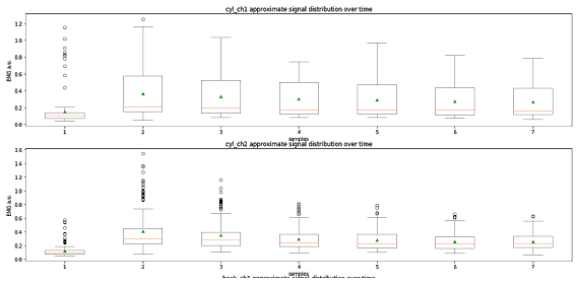

Figure 4: Distribution of cylinder grip signals channels 1 & 2 over time (median: orange line, mean: green triangle).

Rather than investigate how the signals might deviate from the grip baseline, the hypothesis that the data can be considered to be a representative sample by which a reliable baseline can be calculated must be evaluated. The criteria to accept this hypothesis is:

1. The signals grouped by grip and channel must be statistically different.
2. The signals for subjects within a grip-channel grouping must be ***not*** be statistically different..

1) Criteria 1

The signals grouped by grip-channel are not normally distributed; in fact in several cases there appeared to be more than one population within the groupings (*Figure 5*).

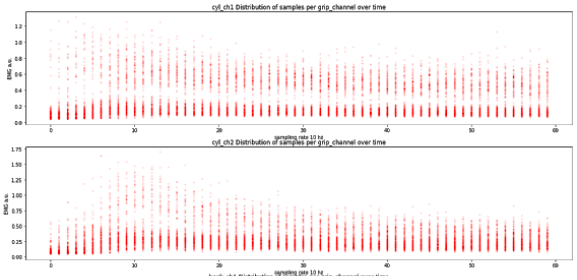

Figure 5: Distribution density of cylinder grip channels 1 & 2 over time.

This observation led to a choice to use a Kruskal Wallis H-test, or one-way ANOVA on ranks, which is a non-parametric method for testing whether samples originate from the same distribution by comparing the median values. To test if the signals by grip-channel are each populations that are statistically different from each other, the null hypothesis must be that the medians of each group of signals by grip are statistically similar. A significant Kruskal–Wallis test indicates that the samples come from different populations, thus rejecting the null hypothesis. A Kruskal Wallis H-test that moves over time with the samples will reveal where samples per grip-channel might be considered part of the same population. Since there are two populations to be tested (6 grips within channel 1 and 6 grips within channel 2), this test is run twice, each test comparing 6 populations.

1) Criteria 2

The signals within each grip-channel grouped by subject appear to adhere to the central limit theorem, due to the fact that the mean of the signals seems to trend in the center of the interquartile range, which indicates that the samples are consistently normally distributed (*Figure 6*).

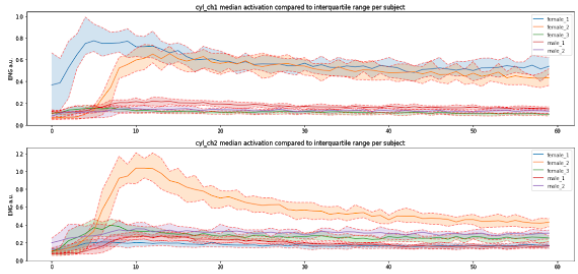

Figure 6: Median compared to interquartile range of cylinder grip channels 1 & 2 over time.

However, the standard deviations are varying between the distributions, meaning a Kruskal Wallis H-test is again the appropriate choice. In this situation the goal is to prove that signals grouped by subject within each grip-channel grouping are actually from the same population. In other words, in testing the samples, the goal is to accept the null hypothesis: that each subjects signals within a grip grouping are not statistically different and are indeed from the same population. Since there are 5 subjects within each grip-grouping, this test must be run on 12 populations (6 groups of 5 subjects for channel 1 and the same for channel 2) for a total of 12 tests, each comparing 5 populations.

V. RESULTS

Computing a confidence interval for the mean of each signal over time shows that after initial activation, the mean signal seems to stabilize within a range of about 0.2 activity units. With the the addition of channel 2, the mean signals are sufficiently distinct from one another (*Figure 7*).

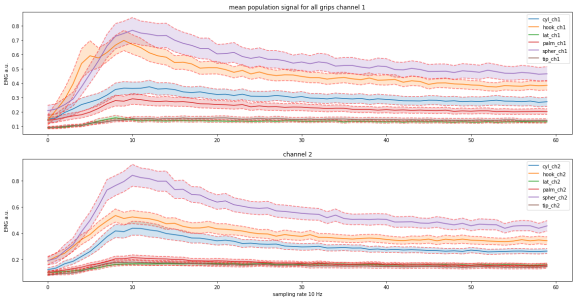

Figure 7: Mean of all signals per grip for channels 1 & 2 with 95% confidence interval.

However, this baseline may not be useful in a practical sense if the average signal doesn't accurately describe the variance within the population per grip signal, thus it is time to examine the results of the *Criteria 1* moving Kruskal Wallis H-test, which is testing the null hypothesis that the signals grouped by grip and channel are all from the same distribution. The results overtime showed a consistent *p* value <0.001 for both channel tests (each channel containing 6 grip populations) (*Figure 8*).

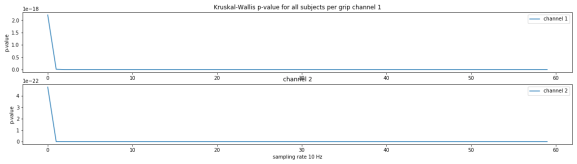

Figure 8: Moving H-test p value for Criteria 1.

The *Criteria 2* test was necessary to test the hypothesis that subjects within a grip-channel grouping are from the same distribution. Again the moving Kruskal Wallis H-test shows a *p* value <0.01 consistently for all 12 groups of 5 subjects within channel 1 (with the exception of the first second of the hook grip grouping) and a *p* value <0.002 for all 12 groups of 5 subjects within channel 2 over time (*Figure 9*).

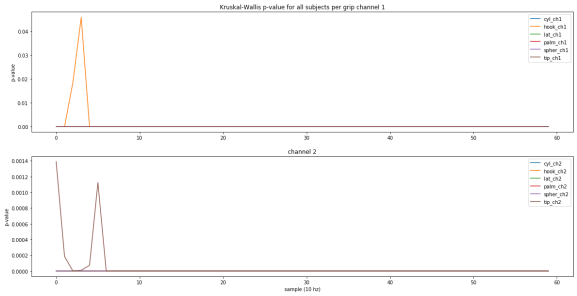

Figure 9: Moving H-test p value for Criteria 2.

VI. DISCUSSION

In reference to the exploratory data analysis, it is clear that each extracted feature has different advantages over one another. The signal envelope, while an ideal candidate for capturing the most variety in each signal of the populations, was very high in samples over time compared to the mean amplitude extracted over adjacent windows, which captured the same amount of variation with many less samples over time (6 seconds at 500 Hz comes out to 3000, vs. 10 Hz and only 60 samples). This makes the mean amplitude extracted over adjacent windows a more attractive candidate for calculating a baseline. In contrast, the relative smoothness of the mean amplitude extracted over overlapping windows helps reduce signal variation substantially, making it a better choice for new signals being compared to the baseline signal calculated using the mean amplitude over adjacent windows.

After computing the mean signal for each grip per channel, the minimally overlapping confidence intervals make it reasonable to say that a combination of the mean of both channels makes a distinct description of each grip signal-wise for all subjects (with an exception for both channels of tip and lat grips) (*Figure 7*).

However, this all means very little if the data being analyzed here cannot be considered to be a representative sample of a larger population, since that would make any baseline calculated here fairly unreliable in terms of generalizing to new populations.

The *Criteria 1* Kruskal Wallis H-test indicates that the null hypothesis (that the signals grouped by grip are from the same distribution) should be rejected. This is ideal for this analysis, as it indicates that the signals collected from one hand gesture are distinct from signals collected from another hand gesture.

The next question that needs answering is if it matters who the subject the signal is collected when within the same grip-grouping. The *Criteria 2* Kruskal Wallis H-test addresses this question directly. The results indicate that the null hypothesis (that the subjects within the grip-grouping are from the same distribution) should be rejected once again. This indicates that within each grip-grouping per channel, signals from one subject are distinct from

signals collected from another subject. To answer the above question, according to this dataset, it does matter which subject the signals are being recorded from.

Recall that it was necessary for *Criteria 1* to reject the null hypothesis, but for *Criteria 2* to accept the null hypothesis. The results found here show that while the signals are distinct when compared between grips by channel, but are not statistically similar when the signals between subjects within those grip-groupings. If the population here was to be considered a "representative" sample (i.e. the results could generalize to newer data) one might expect that the subjects within the grip-groupings would be of the same distribution, but **the results here seem to suggest that this data cannot be reasonably representative of a larger population without further descriptive variables to describe the data more effectively.**

So clearly something is causing these samples to be part of different distributions. Differences between sex of the subject collected seemed random at best. Recall however, that the speed and force were intentionally left to the subject's will by the data collectors. These are some unobserved variables that could be contributing to the multiple subject distributions within the grip-groupings, but are impossible to account for with the quality of data available in this dataset.

VII. CONCLUSION

The unfortunate conclusion of this analysis is that without a higher feature dataset, calculating a practical baseline utilizing the methods outlined above is really not possible. The data-collection method by which this dataset was composed was originally intended to be proof-of-concept for the effectiveness of simplistic EMG signal collection, which made the data-collection ideal for use in a therapeutic setting where practicality and ease of implementation are valuable attributes.

While this data set was not quite enough to meet the needs of this project, that is not to say that higher feature count is absolutely necessary. With only 5 subjects, there is little proof that the results of the tests above could not change substantially in the presence of more EMG signals collected from a larger population of test subjects. In the context of this analysis alone, a higher feature count would be absolutely necessary, but with more data from more subjects there is not necessarily a need for a data-collection method more sophisticated than the one by which the data for this project was collected.

In consideration for new therapeutic documentation processes, this concept should remain a viable consideration for further testing for those willing to collect the data (or spend the money).