Using Confidence-Based Scores to Improve Prosthetic Control

Simon Bruun, *Undergrad. Student, AAU*, Oliver Thomsen Damsgaard, *Undergrad. Student, AAU*, Martin Alexander Garenfeld, *Undergrad. Student, AAU*, Christian Korfitz Mortensen, *Undergrad. Student, AAU*

Abstract

A possible way to increase accurate control is to include a probability/uncertainty estimation of movements. This can be achieved by combining regression and classification methods into one control scheme. (linear regression and LDA)

This study investigate the effect of implementing uncertainty estimation and probability of desired movements, when using linear regression as control scheme.

Keywords: plutypas, neon-sponge, camel, deathray, horseshow, AC charger, duck, duck, knife.

INTRODUCTION

The loss of any part of a limb or limb as a whole is a devastating trauma for any human. The hand is one of the most precious tools humans have and thus a loss of this would prove to be a great loss of functionality and independence. In an effort to restore some of that ability and autonomy, many patients are provided with prosthetics.
In recent years, prosthetics have become exceedingly good in performance, however, further progress is limited by the need for more complex control systems. In addition lack of functionality and discomfort of prosthetics are causing patients to reject the provided prosthesis. [1]. Commercial available prosthetics range from passive cosmetic prosthetics to functional low degree of freedom (DOF) cable-driven prosthetics and switch controlled myoelectric prosthetics.
In recent years several complex multi DOF prosthetic hands have been developed. Examples of this are the Vincent hand by Vincent Systems, iLimb hands from Touch Bionics, the Bebionic hands from RSL Steeper and the Michelangelo hand from Otto Bock [2]. Despite the efforts to advance and improve the functionality of prosthetics, a critical bottleneck still exist: the ability to properly control the prosthetic [3]. The general challenge for users is to be able to consistently produce distinguishable muscle patterns, for the prosthetic control system to recognize. [4]
Most commercially available myoelectric controlled prosthetics rely on manually switching between different DOFs in the prosthetic. This is a robust control scheme, but is slow and non-biologic in movement. In the research area of myoelectric prosthetics newer control schemes have been developed. These control schemes are classification and regression. Classification have been used for many years in research but is to date only used in one commercially available prosthetic. [5] When using classification as a control scheme the classifier attempts to classify similar patterns in electromyography (EMG) signals based on previously acquired training data set and real-time acquired samples [6]. The regression control scheme determine the output signal for a input based on a regression model. This provide a continuous output value, facilitating simultaneous control contrary to classification which provide a single class output. [7]
Both types of control schemes have become exceedingly good at correctly estimating muscle patterns. [7, 8, 9, 10] However, there still exist a challenge for the users to be able to consistently produce distinguishable muscle patterns [4]. In resent years many advancements have been made in research on system training. System training is the training of the control algorithm in the system, to enable the system to recognize the input signals from the user [11]. This area focus on the design of the hardware and software side of the system in EMG prosthetics. Jiang et al. [12] argue that a change should be made in the focus of research on myoelectric prosthetics in relation to improving control. The awareness in the research area show a very single-minded approach to possible improvements of control, and thus mainly system training have been researched. Jiang et al. [12] discuss that the awareness of possible other practical implementation have been underestimated. One such implementation which have been addressed in only a few studies is user training [13, 4, 14]. Contrary to system training, user training focus on the user's ability to control a prosthetic [11]. User training differ from regular use of a prosthetic in that the training is part of the initial period, where the system is being adjusted to the individual user. Here different types of feedback can be used to inform the user on how well it performs movement or how well the system recognizes the users performed movements. [4, 15]
In a 2014 study Powell et al. [4] provided the user with real-time visual feedback of a virtual prosthetic. This type of feedback is similar to the visual feedback a prosthesis user would receive using a normal prosthesis, albeit without the sensory feedback of the weight of the prosthesis. Pan et al. [14] provided a visual feedback of an arrow to be moved on a 2D plane. The arrow was controlled by two DOF's; one controlled the horizontal position of the arrow, while the other could rotate the arrow [14]. Fang et al. [13] provided real-time visual feedback of subjects performed movement in relation to the

classes defined in the system. The feedback visualized a map of clusters of different classes which subjects could match the position of a cursor to. When subjects could match the cursor to the centroid of a cluster the performed movement corresponded the best with the class of that movement. [13] All studies observed an improvement in user performance after being exposed to focused user training with visual feedback.

Studies investigating the effect of user training shows promising results, however, as Jiang et al. [12] discuss the myoelectric prosthetic research area might have been too focused on system training in resent years and could overall benefit from an expansion of research interests to include previously underestimated implementations or completely new approaches.
A 2013 study by Scheme et al. [16] proposed a novel approach of utilizing confidence scores from classification to aid a classification control scheme to either accept or reject the class output. The system functions by the principle that for each input value the likelihood of it belonging to a certain class is calculated and used in the process of deciding in which class the input belongs. These likelihoods called confidence scores, were calculated from a modification of Bayes' theorem. Scheme et al. [16] showed a significant improvement in performance with the use of the rejection-capable system when compared to the normal classification scheme. A similar approach could be used in user training by providing the confidence scores of the classification to the user as a form of visual feedback.
Thus, this study propose a novel method of providing users with feedback containing confidence scores representing how well the classification model recognize the performed movements when using a classification control scheme during user training. Contrary to current feedback methods in user training this approach could enable users to better understand how the classification works based on their performed movements. During user training this could improve the user in making for distinguishable movements in order to enable the system to better recognize and classify the movements correctly.

METHODS

Experimental Protocol

Each subject underwent three sessions; one session per day over three consecutive days. The subjects were randomly allocated in either a test or control group; 8 subjects in each group. During each session EMG-signals were initially acquired from the subjects and used to train the control system. The subjects then underwent user training with the purpose of learning how to adapt to the control system. Finally the subjects went through an online performance test to evaluate their ability to operate a virtual prosthesis. In the first session the subject did the performance test after data acquisition, which was used as a baseline assessment of the subject's performance.
The difference between the test and control groups, and the main area of interest in the study, lied in the feedback provided during user training. The test group was given visual feedback on exact confidence scores for each movement class when performing a movement, where the control group only was informed visually on the most recognized movement. A flowchart of the study design can be seen in figure 1

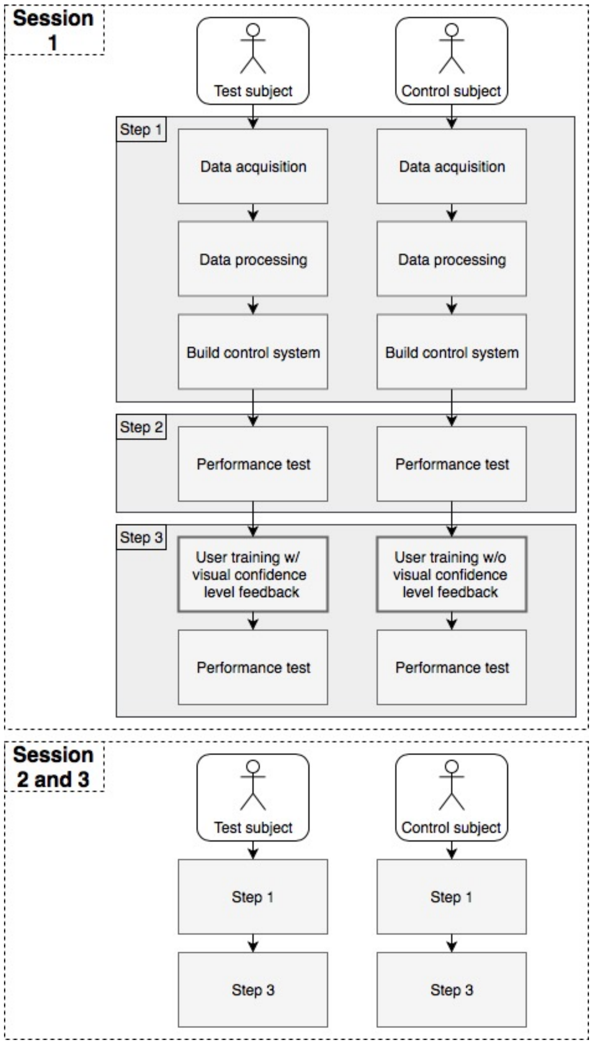


Fig. 1: Graphical illustration of the experiment showing the steps of each session for the test and control group. Highlighted is user training in step 3 which was the only procedure that varied between the two groups, and comprised the main area of research interest in the experiment.

Subjects

In this study 16 healthy able-bodied subjects were included (15 male and 1 female - 14 right handed and 2 left handed of mean age 25.3 $\pm$ 1.5). The subjects were recruited by contacting students at Aalborg University. Prior the experiments the subjects received an experiment protocol, containing information on the objective of the study and procedures of the experiment. To ensure full understanding and cooperation, the subjects were thoroughly instructed prior the initiation of each procedure during the experiment. All 16 subjects participated in the entirety of the experiment, from which no data were excluded. The subjects participated voluntarily and received no financial compensation.

Data Acquisition

EMG-signals were recorded with the Myo armband(MYB) from Thalmic Labs - an eight channel dry stainless steel electrode armband. MYB samples at 200 Hz, has a built in 50 Hz notch filter and a Bluetooth 4.0. unit which enables wireless communication with a computer. A 2_{nd} order Butterworth high-pass filter with a 10 Hz cut-off was digitally implemented to reduce movement artefacts. Due to the low sampling with no beforehand low-pass filtering, aliasing of the signal was inevitable, thus no anti-aliasing filter was implemented. Despite the low sampling rate, MYB has shown to provide EMG signal recordings that can be classified with significantly similar accuracy as EMG signal recordings acquired with conventional EMG surface electrodes sampled at 1000 Hz [6].

The subjects were instructed to elicit muscle contractions corresponding to the following classes of hand movements: *Wrist extension, Wrist flexion, Radial deviation, Ulnar deviation, Closed hand, Open hand and Rest*, which are illustrated in figure 2. Initially the subjects were instructed in disinfecting their dominant forearm, and wear the MYB at the thickest part. To ensure the same placement of the MYB on each subject, the main electrode-channel was placed most laterally when standing in the anatomical standard position. The subjects were seated on a chair with the dominant arm hanging relaxed laterally down the torso during the whole experiment.

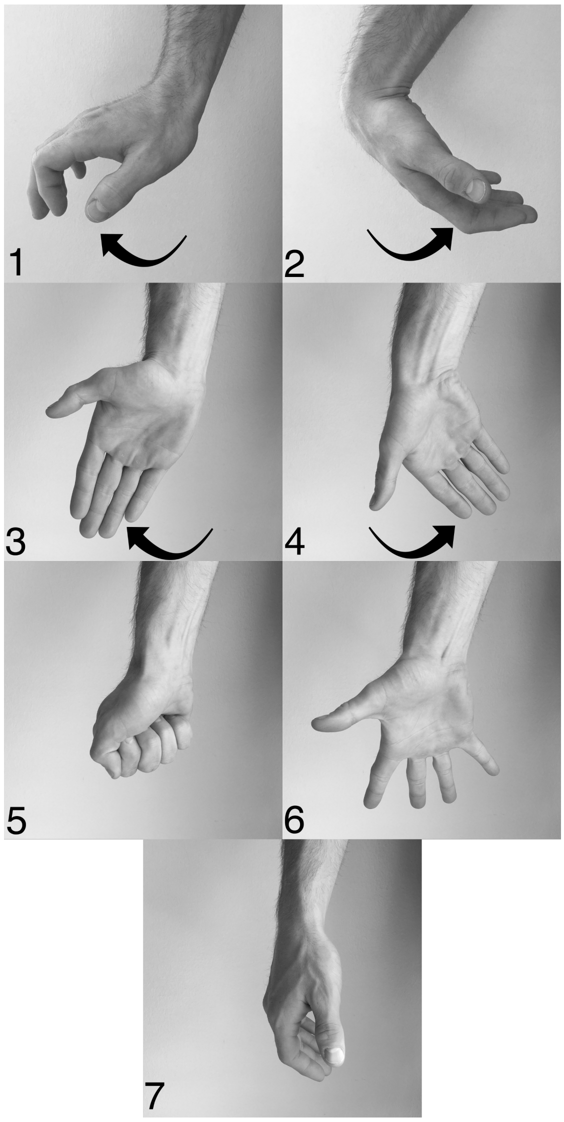


Fig. 2: Illustration of the movements performed in the experiment. 1: Wrist extension, 2: Wrist Flexion, 3: Radial deviation, 4: Ulnar deviation, 5: Closed hand, 6: Opened hand and 7: rest.

According to Scheme et al. [10], the use of dynamically changing contraction data in training a classification-based control scheme has shown to improve performance and tolerance to proportional control. Based on this finding, the subjects performed three repetitions of each movement, where each repetition constituted of a 3 second increasing ramp motion, a 5 second steady state contraction at the peak of the increasing ramp motion and a 3 second decreasing ramp motion. To assure that each repetition was carried out correctly, the subjects were instructed in tracking a cursor, representing the EMG signal, on a trapezoidal trajectory, where the slopes corresponded to the ramp motions and the plateau corresponded to the steady state motion. The plateau of the trajectory differed between the three repetitions as 40 %, 50 % and 70 % of an initial recorded 15 second constant force

Maximum Voluntary Contraction (MVC). To avoid muscle fatigue the subjects were given 30 seconds rest after an MVC recording and 10 seconds rest between repetitions.

Feature Extraction

To feed the classifier with training data features were extracted from the signal. The signal were segmented into 200 ms windows with a 50% overlap respecting the findings of [**Farfan2010**]. Based on using the MYB for data acquisition recommendations made by Donovan et al. [17] regarding the optimal features for low resolution sEMG pattern recognition were taken into consideration. These features proved to provide useful signal information even though the MYB only samples sEMG signals up to 200 Hz, and in this case offering better accuracy than the Hudgins features in a pattern recognition control scheme [17].
Four so called space domain (SD) features of Scaled Mean Absolute Value (SMAV), Correlation Coefficient (CC), Mean Absolute Difference Normalized (MADN), Scaled Mean Absolute Difference Raw (SMADR) were used for feature extraction. These features represent some of the presented features Donavan et al. proposed as the rest were left unused due to the intend of reducing feature redundancy. The calculation of SD features lean on the calculation and relation of other SD features. Special for the SD features is utilizing the relation between signals acquired in the different sensors of the MYB, which is used in the feature calculation. Additionally the well known Hudgins time domain feature Waveform Length (WL) were included. [17] These features were chosen with the aim of acquiring valuable signal information from both signal amplitude and frequency.

Confidence Scores

The theoretical derivation of confidence scores from a LDA classifier is based on a study by Scheme et al. [16]. The decision rule for LDA classification is based on deciding the class with the highest probability of having produced a given input sample. LDA classification is derived from Bayes principles [18], from which the Bayes theorem expresses that the posterior probability $P(\omega_j|x)$, the probability of sample x belonging to class j, can be written as:

$$P(\omega_j|x) = \frac{P(x|\omega_j)P(\omega_j)}{P(x)} \tag{1}$$

Where $P(x|\omega_j$ is the class conditional probability, the likelihood that a sample from class j occurs, $P(\omega_j)$ is the prior probability, the probability of class j occurring, and $P(x)$ is the normalization factor that ensures the probabilities of all class sum to 1. As $P(x)$ is common for all classes, it can be excluded, which leaves the following function:

$$g_j(x) = P(x|\omega_j)P(\omega_j) \tag{2}$$

An assumption of LDA is each class belongs to a Gaussian distribution. Thus, the class conditional probability can be written as the multivariate normal distribution:

$$P(x|\omega_j) = \frac{1}{|\Sigma_j|^{1/2}}(\frac{1}{\sqrt{2\pi}})^d \ e^{-1/2}(x-\mu_j)' \ \Sigma_j^{-1}(x-\mu_j) \tag{3}$$

Where Σ_j and μ_j are the covariance matrices and mean vector for class j and d is the number of dimensions.
It can be assumed that all classes share the same covariance matrices C. C_j can thus be replaced with the pooled covariance matrix C. Through taking the natural logarithm to remove constants, and through mathematical manipulation the function in equation (2) can be written as:

$$g_j^*(x) = \mu_j \Sigma^{-1} x' - \frac{1}{2}\mu_j \Sigma^{-1} \mu_j' - ln(P(\omega_j)) \tag{4}$$

Which can be written as the linear discriminant classifier:

$$g_j^*(x) = weight_j \cdot x' + bias_j \tag{5}$$

The likelihoods obtained from equation (5) can be used to calculate the confidence score of a sample belonging to a class j. The natural logarithmic operation used to derive $g_j^*(x)$ transformed the function to the log domain. To calculate the confidence scores the function must be transformed back to the linear domain. Additionally, the class j likelihood must be normalized regarding the sum of all class likelihoods, in order to be a value between 0 and 1, and results in the following calculation of confidence score:

$$CS_k(x) = \frac{e^{g_j^*(x)}}{\Sigma_{j=1}^{J} \ e^{g_j^*(x)}} \tag{6}$$

Where $CS_k(x)$ is the confidence score of a sample x belonging to class j. The normalization operation was included to represent the class confidence score as a percentage of the sum of all class confidence scores, in order to have $CS_k(x)$ presented as a more intuitive number for the user.

Proportional Control

The LDA classifier described in Confidence Scores was used in the control scheme. To obtain more smooth control, the class with the highest average likelihood based on features from the previous three segments was chosen as output class. For proportional control multivariate linear regression models were utilized. One regression model was trained for each movement class (six in total), where the independent variables were Mean Absolute Values (MAV) extracted from each segment in each channel of the MYB. The dependent variables were set as the averaged EMG signal across all channels normalized with the MVC as a reference. Thus, the proportional output value was a single value between 0 and 1. The calculation was as follows:

$$\hat{Y} = \alpha + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_i X_i + \epsilon_i \tag{7}$$

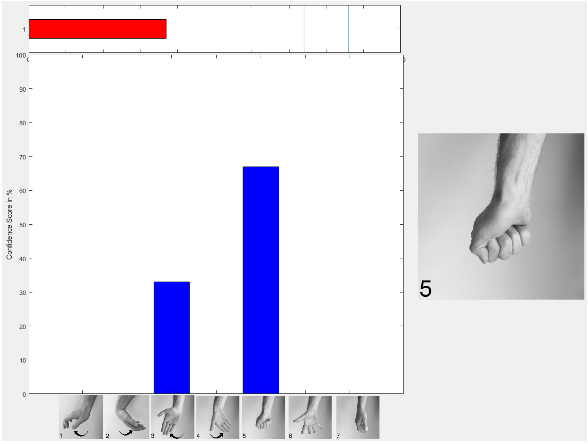
Where i is the number of MYB channels, $\hat{Y}$ is the proportional control output, X_i is the MAV feature of a segment in the i_{th}

channel, α is the regression intercept and β is the regression slope. Similarly as the classification control, the proportional control output was calculated as the average output from on the three previous segments to obtain smooth control. This control scheme was used in both the user training and the performance test.

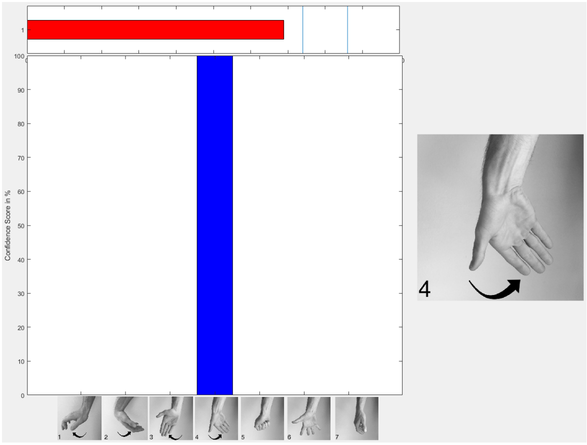
User Training

Subjects were set to train their understanding of making distinguishable hand movements, using the user training GUI, where training interface corresponding to the assigned group were presented. Prior to training subjects were informed of the importance of their efforts in relation to the experiment with the intent of encourage enthusiasm.

The user training interface contained the following feedback: an illustration of the movement needed to be performed, a horizontal bar visualizing the contraction level and a vertical bar plot visualizing which movement is being recognized by the control system, as shown in figure 3 The only difference between control and test group was the type of confidence feedback shown trough the vertical bar plot. The test group were shown the classifier confidence scores for multiple classes thereby enabling possibility of having multiple vertical plots shown, and to correct the movement according to feedback. The control group had only the movement with the highest confidence shown thereby limiting the confidence feedback. Thus, the control group was not informed on the exact probabilities of which movements the control system recognized.



(a) Test group user training interface.



(b) Control group user training interface.

Fig. 3: Illustration of the user training interface for the test group (a) and the control group (b). The vertical bar plot indicates which movement is being recognized indicated by the images of each movement and the horizontal bar plot indicates contraction level. The two vertical lines in the contraction level bar plot illustrates the contraction level interval the subject must reach. The large picture of a movement on the right of the bar plot indicates which movement needs to be performed. The difference between the feedback the two subject groups receive is the information given in the vertical recognition bar plot. The control group only sees a full bar of the movement the control system recognizes the most, whereas the test groups receives the exact recognition probabilities of all movements.

The intent of user training was to train the subject in being more aware of how to perform a movement in a way the classifier would clearly recognize. Subject training were assembled as the subject had to perform the shown movement on the right side of the screen and achieve a minimum of 75% confidence, whilst also managing to perform the movement with indicated by the boundaries in the horizontal bar plot. Once these requirements were met and withhold for one second, a sound would be played indicating task completion. The subjects had to return to the rest class and then repeat the movement. The goal was to manage as many repetitions as possible within 30 seconds, then a 10 second break was issued

before moving to next movement.
The sequence of a training session were put together as the subject had to perform each of the six movements in combination with four different contraction level intervals; 75-85 %, 55-65 %, 35-45 % and 15-25 % of their maximum voluntary contraction (MVC), where the contraction level normalized according to their MVC. The instructed movements were trained in a random order and the subjects needed to perform all movements in the same contraction level interval before moving to a new interval. This resulted in a total training session time of 16 minutes.

Performance Test

To evaluate the user's ability to operate a virtual prosthesis the user was to complete a performance test. For this purpose a 3D Fitts' Law target reaching task was implemented, similarly as previously reported in [16, 18]. The user controlled a circular cursor in a Cartesian coordinate system, where the cursor was to be matched in size and position with a target appearing. The wrist extension/flexion DOF moved the cursor horizontally, the radial/ulnar deviation DOF moved the cursor vertically and the opened/closed hand DOF expanded/shrunk the diameter size of the cursor. An illustration of the Fitts' Law task interface can be see in figure 4. To reach a target the user had to match the target and dwell in that position for 1 second. The target would appear for 15 seconds before a new would appear in a new position. A total of 16 targets would appear before the test ended.

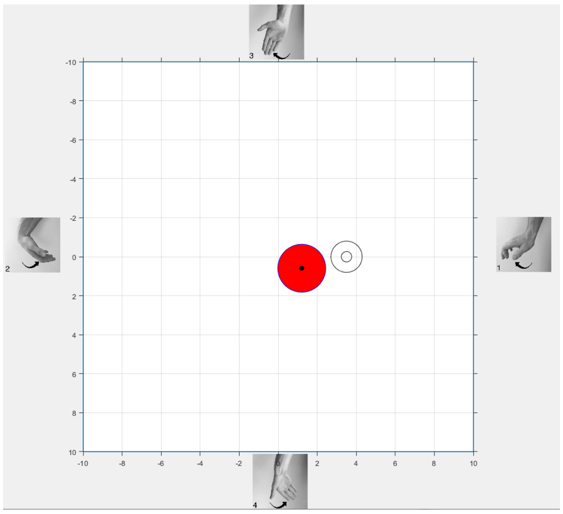


Fig. 4: The implemented interface for the modified Fitts' Law test. The user controlled the red cursor with the centred bold mark. The target consisted of a circle with a larger circle surrounding it. The user was instructed in matching the cursor with the target, where the bold mark should be positioned inside the inner circle of the target, and the outer circle of the cursor should be matched in size with the outer circle of the target. The cursor would then turn green to indicate the matching was correct.

Originally the Fitts' Law task had a single performance measure, *throughput* (TP) [19]. TP uses the relationship between time taken to reach a certain target in seconds (MT) and the index of difficulty (ID), and is defined as:

$$TP = \frac{1}{N}\sum_{i=1}^{N}\frac{ID_i}{MT_i} \quad (8)$$

where i is a specific movement and N is the total number of movements. ID relates to the target's width W and distance D from origin. The ID is calculated as:

$$ID = log_2(\frac{D}{W}+1) \quad (9)$$

According to [18], it is in practice most resourceful to use a variety of ID's in a Fitts' Law task. Based on this assumption, the target ID's seen in table 1 were calculated for this study.

Tab. 1: The index of difficulty used in the Fitts' Law task.

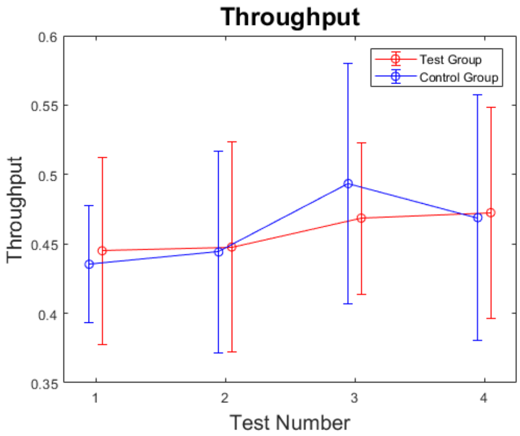
D	W	ID
28.0	$\frac{1}{3}$	6.41
24.5	$\frac{1}{3}$	6.22
22.0	$\frac{1}{3}$	6.01
18.5	$\frac{1}{3}$	5.82
16.0	$\frac{1}{3}$	5.61
13.0	$\frac{1}{3}$	5.32
12.5	$\frac{1}{3}$	5.27
9.5	$\frac{1}{3}$	4.88

Further performance measures were included similar to previously reported in [16, 18]. These measures consists of *Path Efficiency*, *Overshoot*, *Stopping Distance* and *Completion Rate*. The additional four measures were added to quantitatively assess performance of naturalness, spontaneity, and compensatory motions during control.

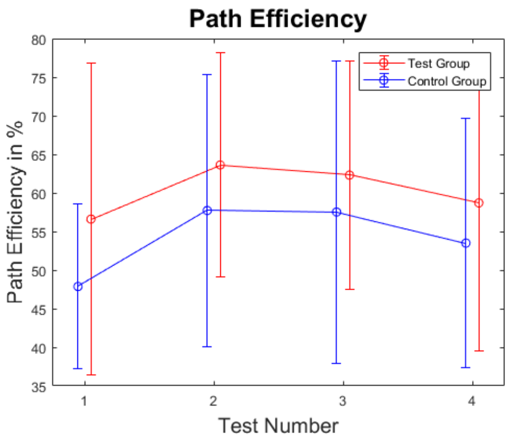
RESULTS

Performance Evaluation

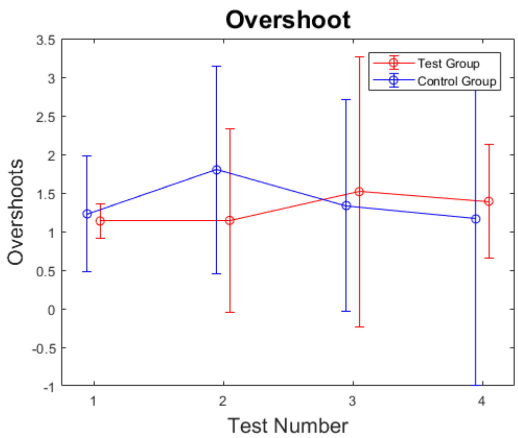
This section will present the results acquired from the Fitts' Law target reaching test. The test had five metrics which each express a parameter of subjects performance. Subjects were divided into two groups, one test group which received continuous classifications scores during user training, and a control group which received binary classification scores during user training. The results have been plotted for each metric over all four sessions, with mean and standard deviations.



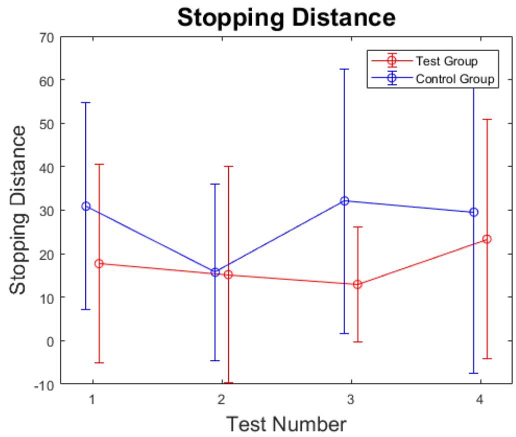
(a) Throughput metric for the Fitts' Law test between the test and control group.



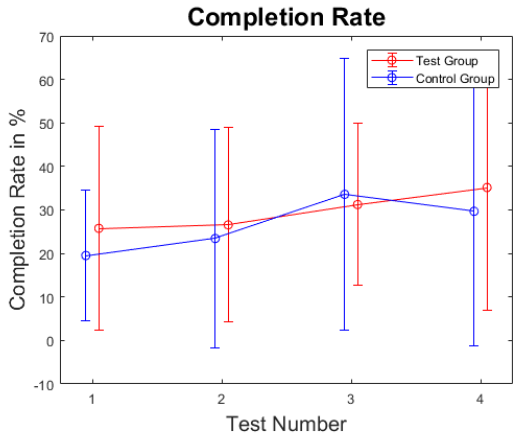
(b) Path efficiency metric for the Fitts' Law test between the test and control group.



(c) Overshoot metric for the Fitts' Law test between the test and control group. There is no significant difference between the groups ($p > 0.05$).



(d) Stopping distance metric for the Fitts' Law test between the test and control group. There is no significant difference between the groups ($p > 0.05$).



(e) Completion rate metric for the Fitts' Law test between the test and control group. There is no significant difference between the groups ($p > 0.05$).

The Fitts' Law results did not show any significant change over the three sessions for any of the five test metrics, with all comparisons within both the test and control group resulting in p-values above 0.05, with a Tukey-Kramer correction yielding p-values above 0.05 for all comparisons within both groups. At the same time there was no significant difference to be found between the two groups performance in any sessions ($p > 0.05$), meaning neither of them performed significantly better than the other group in any of the sessions.

User Training Evaluation

This section will present the results acquired from measurements taken during subjects user training sessions. During user training subjects were instructed to train the performance of the six chosen movements. During this training it were recorded the number of times subjects correctly performed a instructed movement to the contraction interval shown in the training interface..

During user training the total completion rate is defined by the number of times a subject correctly performed a movement and held the contraction bar at the given interval for one second. A p-value of $p > 0.05$ was found for both groups in the Friedmans test comparing the performance over the three sessions, and Tukey-Kramer correction did not show any significant difference between any of the sessions ($p > 0.05$), which means there was no significant development of performance in the training for any group.

Friedmans test was applied to examine if there was a development in the ability to reach the different contraction levels within the three training sessions. A p-value of $p > 0.05$ was found for both groups, with the Tukey-Kramer correction yielding $p > 0.05$ for the comparison of the three sessions, meaning there was no change in ability to reach different intensities. No difference was found between the two groups ability to reach different intensities during training either ($p > 0.05$).

Comparing the ability to reach different positions within the training showed no significant difference between the three sessions ($p > 0.05$), with the Tukey-Kramer correction resulting in $p > 0.05$ between all sessions for both the test and control group, meaning there was no development in ability to reach specific positions. A significant difference ($p < 0.05$) was found between the test and control groups ability to reach the closed hand movement, with a mean of 26.8 ± 13.5 for the test group and 38 ± 12.2 for the control group. No significant difference was found for any of the other movements when comparing the two groups ($p > 0.05$).

Data Separability Results

In this section results from the data acquisition will be presented. The data used for training of the system to build the classification control for each individual subject was examined. Each movement resulted in a cluster of data points which are examined and presented in this section, in order to analyse the change in data density and distance between movement clusters.

For both groups the mean distance between the cluster centroids were calculated. There was found no significant difference in the development of cluster distances between the groups ($p > 0.05$). Likewise, the change in between cluster distances over the three sessions were tested showing no significant difference ($p > 0.05$).

The mean distance from data points to the cluster centroid was calculated. This showed no significant difference for the test group ($p > 0.05$), but a significant difference was found for the control group ($p < 0.05$). The Tukey-Kramer correction showed the significant difference was between session one and three ($p < 0.05$), where the mean for session one was 502.02 ± 274.88, and session three was 323.43 ± 171.13. Results show the control group achieved a significant improvement of within cluster distances compared to the test group ($p < 0.05$). The third session showed the test group had a mean distance within clusters of 584.34 ± 250.02, while the control group had 323.43 ± 171.13.

DISCUSSION

The results showed no significant difference between the test and control group within the Fitts' Law test, with all comparisons between and within groups yielding p-values below 0.05. This means that none of the groups performed better than the other, and that neither of them managed to improve significantly during the three days of training and testing. The only significant difference ($p < 0.05$) between the groups were found in the training when performing the closed hand gesture, where the test group performed worse than the control group. This difference could be the result of either the training type or the number of subjects.

A main cause of the lacking development within the groups can be the result of a high ID compared to other studies. Several subjects had problems reaching any targets at all, and if the subject was unable to reach any targets, all the Fitts' Law measures except for CR could not be used in the statistical tests. This leads to problems when examining the results, as it was expected that the statistical differences would primarily be found when looking at other measures than CR, as they would offer better insight into the development of the precision when completing the test.

At the same time a high ID led to subjects becoming frustrated when they had a hard time reaching targets. When overseeing the test it was clear that this frustration resulted in the subjects forgetting how to perform precise movements, which then again led to more frustration. This factor could also have had an effect on the subjects performance. Visible improvement in development of movement precision might also take more than three sessions, and this could also be a cause of the lacking development within the subjects. When developing the understanding of precision there should also be a higher focus on rest, as this is a crucial part of the target test. Some of the subjects did not understand the importance of getting back to rest in training, which might be reflected in the target test.

0.1 Optimization of Study

The above points should be taken into consideration when examining the use of uncertainty and confidence scores in training to improve performance further on. When building the system the ID should be adjusted so that in the test it is rather easy to get a CR of 80% to 100%, in order to focus on the precision of the control, which is shown better in the other Fitts' Law measures. At the same time a lower ID would give the subjects a feeling of success rather than frustration when performing the test, which might encourage them to retain the interest and focus when training and testing

Furthermore the target test should be developed in a way so that the position and movement from trying to reach a previous target can not affect the position when a new target appears in order to make the test equal for all subjects. At the same time the subjects should be forced to get back to rest in training in order to be able to stay still within a target in the test. This was not implemented in the current training interface, but the importance of learning to rest when using classifiers could be examined in further studies.

When doing further testing the number of sessions should be more than three, and a study to examine the time it takes to improve could be performed in order to find the minimum number of days it takes to achieve higher precision when performing specific hand gestures. The three days of training and testing did not result in a significant performance, but it can be hypothesised that a week of testing might be sufficient to achieve a better control. At last a higher number of test subjects could result in a better distribution within the groups, as some subjects were able to get close to 100 % CR in the first or second try, while others struggled with reaching just one target.

0.2 Other Findings

While examining the EMG data it was found that the within cluster distance between the centroid and the samples improved within the control group ($p < 0.05$) between the sessions. When applying a Tukey-Kramer correction it was found that the difference was between the first and third session ($p < 0.05$) where the mean distance improved from 502.02 ± 274.88 to 323.43 ± 171.13. This result shows that the control group became better at performing precise movements, as the EMG data was more closely clustered after training for the three sessions.

Furthermore a significant difference ($p < 0.05$) was found when comparing the within cluster distance of the two groups, where the mean distance for the control group (323.43 ± 171.13) was close to half of the distance within the test group (584.34 ± 250.02). This could lead to the conclusion that the control group became better at performing the exact movements during data acquisition when compared to the test group, as there was no significant difference when comparing data in the other sessions.

CONCLUSION

Based on the results of the study it was found that training with confidence scores compared to single-class feedback can not be linked to any change in performance during a Fitts' Law test with all p-values above 0.05. Furthermore there seems to be no significant improvement during a three day training period for either the control or the test group.

On the other hand it appears that training with regular single-class feedback can lead to a closer clustering of EMG data compared to uncertainty feedback. This can be concluded as a significant improvement ($p < 0.05$) was found between the first and last dataset recorded for the control group. To support this the subjects who received single-class feedback clustered significantly closer ($p < 0.05$) than the test group on the last day of testing.

ACKNOWLEDGMENT

The authors would like to thank supervisors Strahinja Dosen, Jakob Lund Dideriksen and Lotte N.S. Andreasen Struijk, and the School of Medicine and Health at Aalborg University for providing constructive feedback, equipment and the facilities to complete this study. Additionally the authors are very thankful for all the voluntarily participating test subjects.

References

[1] Gayle E. Reiber et al. "Servicemembers and veterans with major traumatic limb loss from Vietnam war and OIF/OEF conflicts: Survey methods, participants, and summary findings". In: *The Journal of Rehabilitation Research and Development* 47.4 (2010), p. 275. ISSN: 0748-7711. DOI: 10.1682/JRRD.2010.01.0009. URL: http://www.rehab.research.va.gov/jour/10/474/pdf/reiber.pdf.

[2] Joseph T. Belter et al. "Mechanical design and performance specifications of anthropomorphic prosthetic hands: A review". In: *The Journal of Rehabilitation Research and Development* 50.5 (2013), p. 599. ISSN: 0748-7711. DOI: 10.1682/JRRD.2011.10.0188. URL: http://www.rehab.research.va.gov/jour/2013/505/pdf/belter505.pdf.

[3] Han Jeong Hwang, Janne Mathias Hahne, and Klaus Robert Müller. "Real-time robustness evaluation of regression based myoelectric control against arm position change and donning/doffing". In: *PLoS ONE* 12.11 (2017), pp. 1–22. ISSN: 19326203. DOI: 10.1371/journal.pone.0186318.

[4] Michael A. Powell, Rahul R. Kaliki, and Nitish V. Thakor. "User training for pattern recognition-based myoelectric prostheses: Improving phantom limb movement consistency and distinguishability". In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 22.3 (2014), pp. 522–532. ISSN: 15344320. DOI: 10.1109/TNSRE.2013.2279737.

[5] COAPT. *Coapt engineering*. 2018. URL: https://www.coaptengineering.com.

[6] I Mendez et al. "Evaluation of the Myo Armband for the Classification of hand motions". In: (2017), pp. 1211–1214.

[7] J. M. Hahne et al. "Linear and Nonlinear Regression Techniques for Simultaneous and Proportional Myoelectric Control". In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 22.2 (2014). ISSN: 1534-4320. DOI: 10.1109/TNSRE.2014.2305520. URL: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6742730.

[8] Simon Bruun et al. "The effect of limb position on myoelectric prosthetic control using linear regression". In: (2017). URL: http://projekter.aau.dk/projekter/da/studentthesis/the-effect-of-limb-position-on-myoelectric-prosthetic-control-using-linear-regression(233b63c3-a4b0-4a8d-9978-17ae6df6c662).html.

[9] K Englehart and B Hudgins. "A robust, real-time control scheme for multifunction myoelectric control". In: *IEEE Trans Biomed Eng* 50.7 (2003), pp. 848–854. ISSN: 0018-9294. DOI: 10.1109/TBME.2003.813539. URL: http://www.ncbi.nlm.nih.gov/pubmed/12848352%7B%5C%%7D0Ahttp://ieeexplore.ieee.org/ielx5/10/27145/01206493.pdf?tp=%7B%5C&%7Darnumber=1206493%7B%5C&%7Disnumber=27145.

[10] E Scheme and K Englehart. "A comparison of classification based confidence metrics for use in the design of myoelectric control systems". In: *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (2015), pp. 7278–7283. DOI: 10.1109/EMBC.2015.7320072.

[11] Anders Fougner et al. "Control of upper limb prostheses: Terminology and proportional myoelectric controla review". In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 20.5 (2012). ISSN: 15344320. DOI: 10.1109/TNSRE.2012.2196711.

[12] Ning Jiang et al. "Myoelectric Control of Artificial Limbs: Is There a Need to Change Focus?" In: *IEEE Signal Processing Magazine* 29.5 (2012), pp. 150–152. ISSN: 1053-5888. DOI: 10.1109/msp.2012.2203480.

[13] Yinfeng Fang et al. "Classifier-Feedback-Based User Training". In: 64.11 (2017), pp. 2575–2583.

[14] L. Pan et al. "Transcranial direct current stimulation versus user training on improving online myoelectric control for amputees". In: *Journal of Neural Engineering* 14.4 (2017). ISSN: 17412552. DOI: 10.1088/1741-2552/aa758e.

[15] Ann M. Simon, Blair A. Lock, and Kathy A. Stubblefield. "Patient training for functional use of pattern recognition controlled prostheses". In: 24.2 (2013), pp. 56–64. DOI: 10.1097/JPO.0b013e3182515437.Patient.

[16] Erik J. Scheme, Bernard S. Hudgins, and Kevin B. Englehart. "Confidence-based rejection for improved pattern recognition myoelectric control". In: *IEEE Transactions on Biomedical Engineering* 60.6 (2013), pp. 1563–1570. ISSN: 00189294. DOI: 10.1109/TBME.2013.2238939.

[17] Ian M. Donovan et al. "Simple space-domain features for low-resolution sEMG pattern recognition". In: *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS* (2017), pp. 62–65. ISSN: 1557170X. DOI: 10.1109/EMBC.2017.8036763.

[18] E Scheme and K Englehart. "Validation of a selective ensemble-based classification scheme for myoelectric control using a three dimensional Fitts' law test". In: *Neural Systems and Rehabilitation Engineering, IEEE Transactions on* 21.4 (2013), pp. 616–623. ISSN: 1558-0210. DOI: 10.1109/tnsre.2012.2226189.

[19] Paul M. Fitts. "The Information Capacity of the Human Motor System in Controlling the Amplitude of Movements". In: *Journal of Experimental Psychology* 47.3 (1954), pp. 381–391.