Using Confidence Scores in User Training to Improve Users' Ability to Control Upper Limb Prosthetics

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

Abstract

Previous studies have shown user training to be an important part of overcoming the challenge of making transradial upper limb prosthetics more accurate, as the systems depend on the user generating the same distinguishable muscle patterns when using the prosthesis. Different methods have been sought when adapting users to perform specific distinguishable movements. This study aimed to investigate whether confidence score feedback from a LDA classifier during user training could improve user performance in a Fitts' Law test compared to a control group who only received single class feedback. 16 able-bodied subjects were recruited for the study; 8 subjects randomly assigned to each group. Each subject went through a three session experiment; one session per day over three consecutive days. During each session the subject received a 16 minutes user training and went subsequently through a Fitts' Law test to evaluate the performance. The results from the Fitts' Law test showed no significant difference between the two groups ($p > 0.05$) and no improvement over the three sessions for either of the groups ($p > 0.05$). There was no difference in the performance during training as well ($p > 0.05$). A significant improvement in cluster dispersion of EMG signals of separate movement was found in the control group, where the third session resulted in more dense clusters both when compared to the first session of the control group and third session of the test group ($p < 0.05$).

Keywords: surface electromyography, lower arm prosthetics, linear discriminant analysis, user training, confidence scores

INTRODUCTION

The loss of any part of a limb or limb as a whole is disability 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. [1]

In recent years, prosthetics have become increasingly better in performance in a clinical environment, however, further progress is limited by the need for more complex control systems. In addition lack of functionality and discomfort of prosthetics, when used in daily life tasks, are causing patients to reject their provided prosthesis. [2] Commercially available prosthetics range from passive cosmetic prosthetics to functional low degree of freedom (DOF) cable-driven prosthetics and more advanced 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 [3]. Despite the efforts to advance and improve the functionality of prosthetics, a critical bottleneck still exist: the ability to properly control the prosthesis [4]. The general challenge for users is to be able to consistently produce distinguishable muscle patterns, which enables the control system to recognize the performed movements accurately. [5]

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-based. Classification have been used for many years in research, but is to date only used in one commercially available prosthetic [6]. When using classification as a control scheme the classifier attempts to classify similar patterns in electromyography (EMG) signals based on previously acquired training data sets and real-time acquired samples [7]. 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 only a single class output. [8]

Both types of control schemes have improved at correctly estimating muscle patterns. [8, 9, 10, 11] In recent years many advancements have been made in research on system training. System training is the training of the control algorithm to enable the system to recognize the input signals from the user [12]. This area focus on the design of the hardware and software side of the system in EMG prosthetics. Jiang et al. [13] 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. [13] discuss that other methods of improving prosthetic devices have been underestimated. One such implementation which have been addressed in only a few studies is user training [14, 5, 15]. Contrary to system training, user training

focus on the user's ability to control a prosthesis [12]. User training is a focused training of the user in learning to adapt to the control system, before the actual use of the prosthesis in daily life. 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. [5, 16]
In a 2014 study, Powell et al. [5] 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. [15] 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. Fang et al. [14] 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. [14] 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 and as Jiang et al. [13] discuss, the myoelectric prosthetic research area might have been too focused on system training in recent 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. [17] proposed a novel approach of utilizing confidence scores from a Linear Discriminant Analysis (LDA) classifier to aid the control scheme to either accept or reject the class output. The system functioned by the principle that for each input value the likelihood of it belonging to a certain class was calculated and used in the process of deciding in which class the input belonged. These likelihoods called confidence scores, were calculated from a modification of Bayes' theorem. Scheme et al. [17] 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 recognizes the performed movements when using a LDA based 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. Additionally, this proposal of user training could improve the user's ability to produce more distinguishable movements, which would enable the control system to better recognize the performed movement.

This study will due to the presented possibilities in improving user training seek to investigate the use confidence scores as a visual feedback to improve the users' ability to control a transradial prosthesis. This is done under the hypothesis that if exposing subjects to user training, in which confidence levels of movement class recognition as feedback, will show statistically significant improvement in performance in a classification-based myoelectric prosthetic control scheme, when comparing to subjects receiving single-class feedback.

METHODS

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 steps of the experiment. To ensure full understanding and cooperation, the subjects were thoroughly instructed prior the initiation of each step 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 reimbursement.

Experimental Protocol

Each subject underwent three sessions; one session per day over three consecutive days. The subjects were randomly allocated to 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 a real time performance test to evaluate their ability to operate a virtual prosthesis. In the first session the subject completed the performance test prior to user training. This test 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 based on exact confidence scores for each movement class when performing a movement, where the control group only was informed visually on the movement class with the highest probability. A flowchart of the study design can be seen in figure 1.

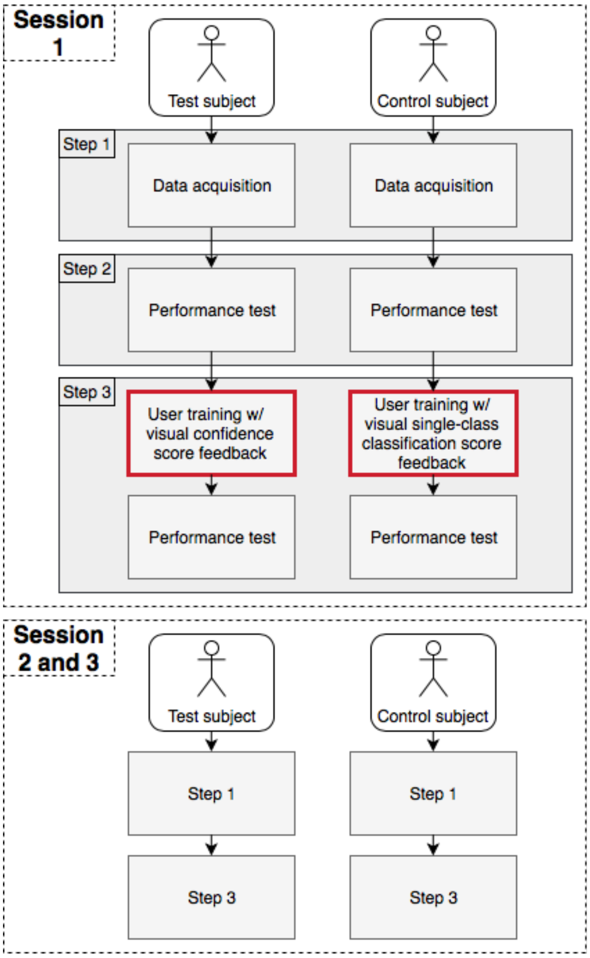


Fig. 1: Graphical illustration of the experiment showing the steps of each session for the test and control group. Highlighted is user training in step 3 which was the only step that varied between the two groups, and comprised the main area of research interest in the experiment.

Data Acquisition

EMG-signals were recorded with the Myo armband (MYB) from Thalmic Labs - an eight channel dry stainless steel electrode armband. The MYB, which 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, the 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 [7].

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. 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.

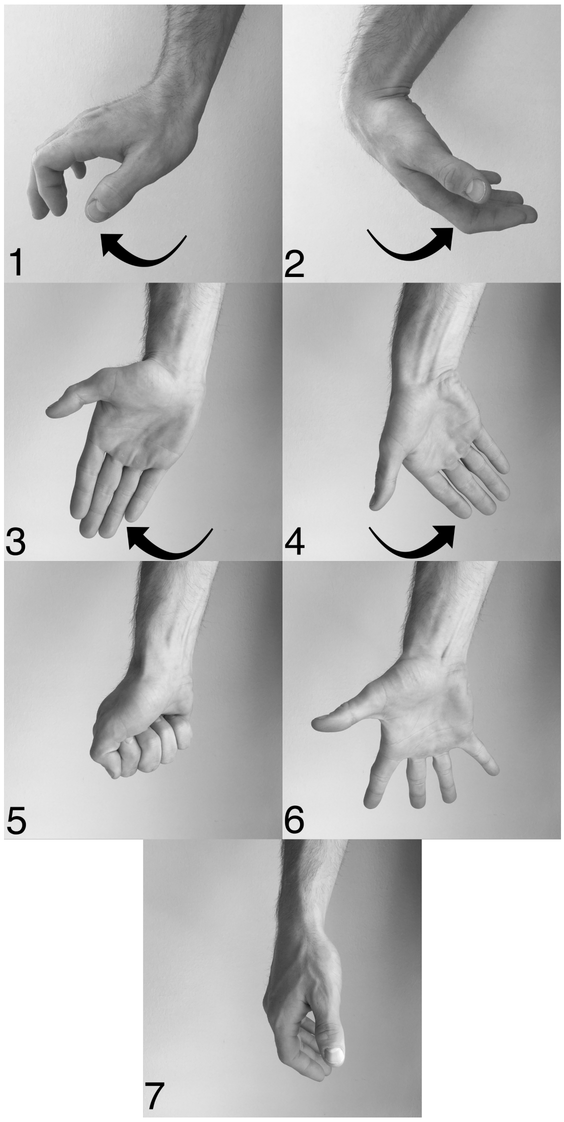


Fig. 2: Illustration of the movements performed in the experiment. 1: Wrist extension, 2: Wrist flexion, 3: Radial deviation, 4: Ulnar deviation, 5: Closed hand, 6: Opened hand, 7: rest.

According to Scheme et al. [11], 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 2.5 second increasing ramp contraction, a 5 second steady state contraction at the peak of the increasing ramp contraction and a 2.5 second decreasing ramp contraction. 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 contractions and the plateau corresponded to the steady state contraction. The plateau of the trajectory differed between the three repetitions as 40 %, 50 % and 70 % of an initial recorded 15 second constant force of 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.

Cluster Dispersion and Separability

The EMG-signal for each movement class acquired from the subjects forms clusters of multidimensional data points. The lower the dispersion of the individual movement class clusters is, the more distinguishable the movements are, and the classifier will recognize the movement classes with higher accuracy. Additionally, a higher distance between cluster centroids will facilitate a higher classification accuracy further. This section describes how to calculate the cluster dispersion and separability of clusters.

To calculate cluster dispersion, the centroid of multidimensional clusters must be calculated as in:

$$C = \left[\frac{\sum_{n=1}^{N} x_n, y_n, \ \dots \ k_n}{N} \right] \tag{1}$$

Where C is the centroid, n is the number of data point in a dimension, N is the total number of data points in a dimension and k is the number of dimensions. To calculate cluster dispersion, the Euclidean distance (ED) from data point p to the corresponding cluster q is computed:

$$ED(p,q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 \ + \ \dots \ + (p_k - q_k)^2} \tag{2}$$

This procedure is performed for all data points in a cluster, from which the average is calculated to obtain a general impression of the cluster dispersion.

To calculate the cluster separability, the ED between cluster centroids is calculated.

Feature Extraction

Before training the classifier, features were extracted from the signal. The raw EMG-signal from each MYB-channel was segmented into 200 ms windows with a 50% overlap respecting the findings of Farfán et al. [18]. Based on using the MYB for data acquisition recommendations made by Donovan et al. [19] regarding the optimal features for low bandwidth sEMG pattern recognition were taken into consideration. These features proved to provide useful signal information even though the MYB only samples sEMG signals with 200 Hz, and in this case offering better accuracy than the Hudgins features [20] in a LDA based control scheme [19].

Four 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 features Donavan et al. [19] proposed as the rest were left unused due to the intent 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 channels of the MYB. Additionally the well known Hudgins time domain feature Waveform Length (WL) was included to cover complexity information in the time domain [21].

Confidence Scores

The theoretical derivation of confidence scores from a LDA classifier is based on a study by Scheme et al. [17]. 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 [22], 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:

$$P(\omega_j|x) = \frac{P(x|\omega_j)P(\omega_j)}{P(x)} \tag{3}$$

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:

$$g_j(x) = P(x|\omega_j)P(\omega_j) \tag{4}$$

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:

$$P(x|\omega_j) = \frac{1}{|\Sigma_j|^{1/2}} (\frac{1}{\sqrt{2\pi}})^d \ e^{-1/2}(x-\mu_j)' \ \Sigma_j^{-1}(x-\mu_j) \tag{5}$$

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 (4) can be written as:

$$g_j^*(x) = \mu_j \Sigma^{-1} x' - \frac{1}{2}\mu_j \Sigma^{-1} \mu_j' - ln(P(\omega_j)) \quad (6)$$

Which can be written as the linear discriminant classifier:

$$g_j^*(x) = weight_j \cdot x' + bias_j \quad (7)$$

The likelihoods obtained from equation (7) 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:

$$CS_k(x) = \frac{e^{g_j^*(x)}}{\sum_{j=1}^{J} e^{g_j^*(x)}} \quad (8)$$

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 the previous section 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:

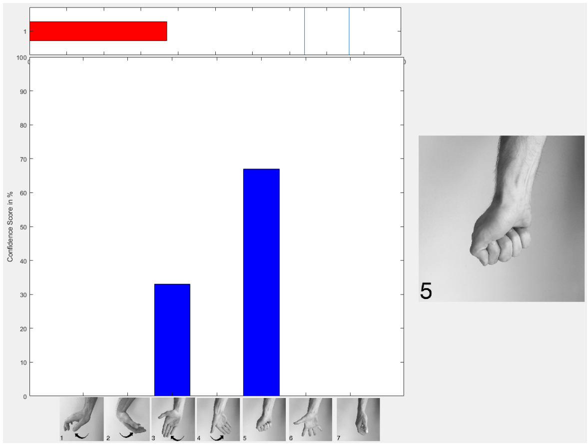
$$\hat{Y} = \alpha + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_i X_i + \varepsilon_i \quad (9)$$

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. ε_i is the error term. 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.

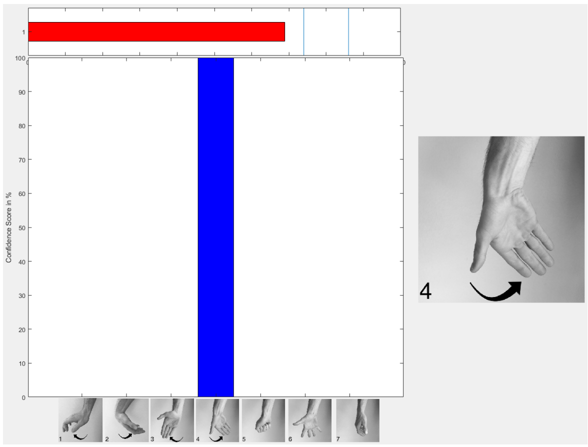
User Training

Subjects were set to train their understanding of making distinguishable hand movements, using a user training interface, where feedback corresponding to the assigned group was presented. Prior to training subjects were informed of the importance of their efforts in relation to the experiment with the intent of encouraging a focused participation.

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 being recognized by the control system. An illustration of the user training interface is shown in figure 3. The only difference between test and control group was the type of confidence feedback shown trough the vertical bar plot.



(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 visualized by the images of each movement; a full bar corresponds to 100 % recognition confidence. The horizontal bar plot indicates contraction level, where a full bar corresponds to the MVC. 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 test group was shown the classifier confidence scores for multiple classes, which enabled the possibility of having multiple vertical plots shown. Thus, more diverse feedback was presented, which the user could utilize to correct the performed movement. The control group had only the movement with the highest confidence shown, thereby limiting the confidence feedback to only one bar visible at a time. Thus, the control group was not informed on the exact probabilities of which movements the control system recognized.

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 recognize as the movement the user actually performed. To motivate the subject during user training a simple task was implemented in the interface. The subject had to perform the instructed movement and achieve a minimum of 75% confidence, whilst also managing to perform the movement within the contraction level interval indicated by the vertical boundaries in the horizontal bar plot. Once these requirements were met and withheld for one second, a sound would appear 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 the next movement.
The sequence of a training session were put together in form of the subject having 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 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

A performance test was developed to evaluate the users ability to operate a virtual prosthesis. The test was implemented as a 3D Fitts' Law target reaching test, similar to methods reported in [17, 22]. The user controlled a circular cursor in a Cartesian coordinate system, where the cursor was to be matched in size and position with appearing targets. Extension/flexion of the wrist moved the cursor horizontally, radial/ulnar deviation moved the cursor vertically and opened/-closed hand increases/decreases the size of the cursor. An illustration of the Fitts' Law test interface can be see in figure 4. To reach a target the user had to match the size and position and dwell within the area for 1 second. The target would appear for 15 seconds or until it was reached, after which a new target would appear and the cursor position would be reset to origo. A total of 16 targets would appear before the test ended. The sequence of targets appearing was different between all four test session, to avoid bias of subjects remembering the sequence in which targets would appear.

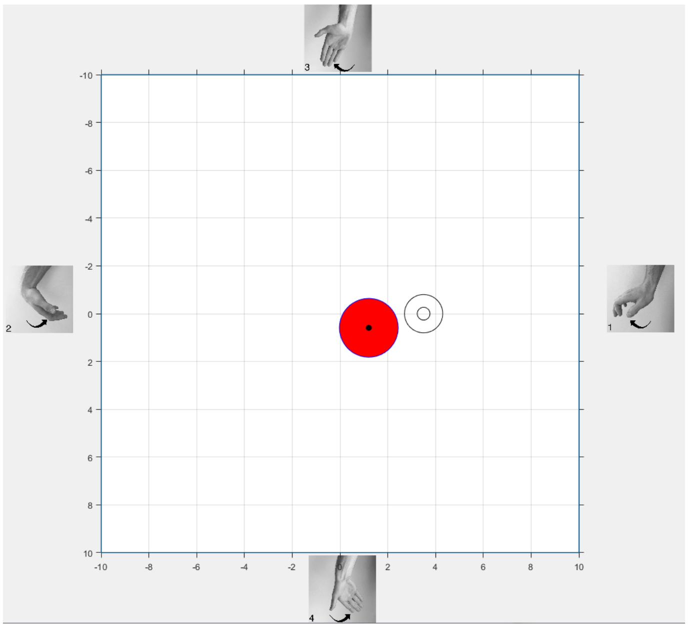


Fig. 4: The implemented interface for the modified Fitts' Law test. The user controlled the red cursor with the centred bold mark. The target consisted of a circle with a larger circle surrounding it. The user was instructed in matching the cursor with the target, where the bold mark should be positioned 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 test had a single performance measure, *throughput* (TP) [23]. 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:

$$TP = \frac{1}{N}\sum_{i=1}^{N}\frac{ID_i}{MT_i} \tag{10}$$

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, where W and D are unitless. The ID is calculated as:

$$ID = log_2(\frac{D}{W}+1) \tag{11}$$

According to [22], it is in practice most resourceful to use a variety of ID's in a Fitts' Law test. 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 test.

Distance	Width	ID
28.0	0.33	6.41
24.5	0.33	6.22
22.0	0.33	6.01
18.5	0.33	5.82
16.0	0.33	5.61
13.0	0.33	5.32
12.5	0.33	5.27
9.5	0.33	4.88

Further performance measures were included similar to previously reported in [17, 22]. 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

Statistics were applied to evaluate improvements in the results obtained in the performance test, user training and data clustering. A Friedmans test was used for multiple comparison and a Tukey-Kramer correction was conducted when detecting an effect. For comparison between groups in each session, a Mann-Whitney U test was applied.

Performance Evaluation

This section presents the results acquired from the Fitts' Law target reaching test. The test had five measures which each expresses a parameter of subjects' performance. Subjects were divided into two groups, one test group which received exact class confidence scores during user training, and a control group which only received a single class confidence score. The plotted mean and standard deviations of each measure in the performance test for session can be seen in figure 5.

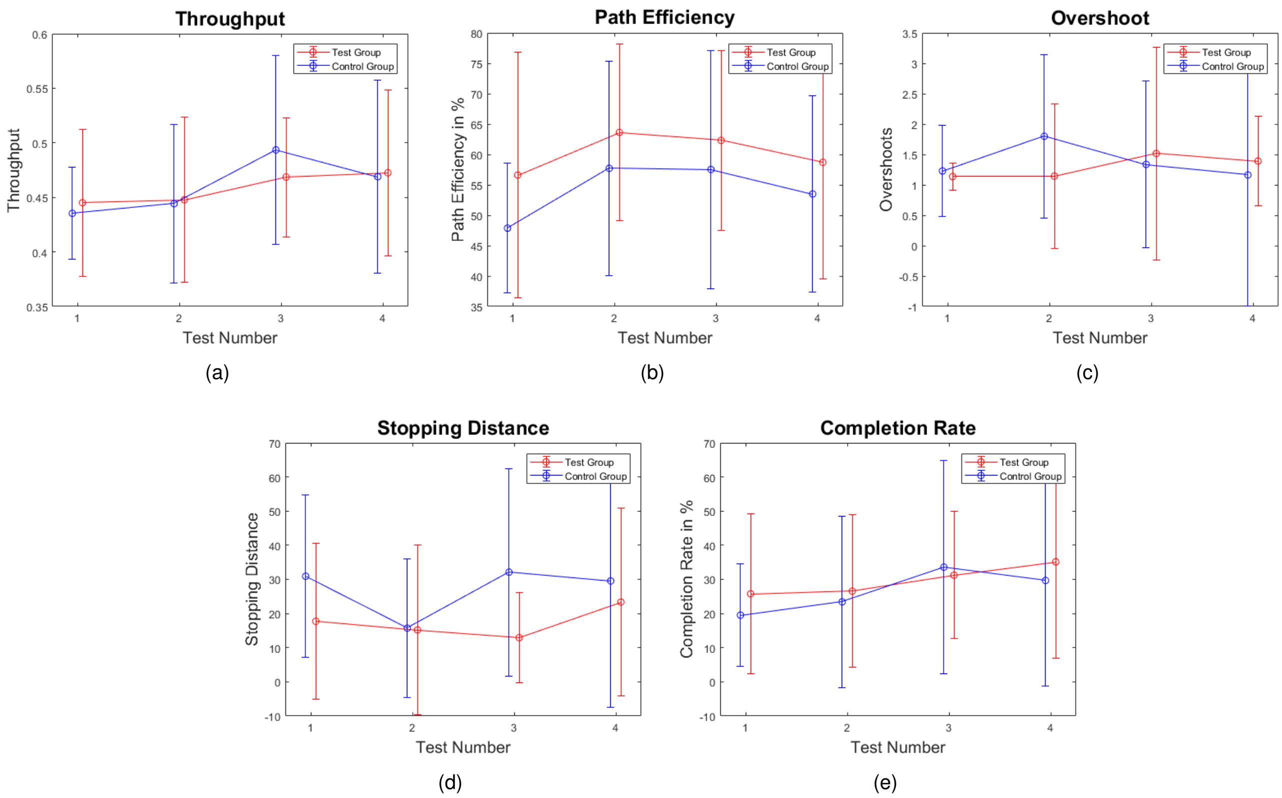


Fig. 5: Figure illustrating the five performance measures; *a) Throughput, b) Path efficiency, c) Overshoot, d) Stopping distance, e) Completion rate*, used for quantifying user performance across all four tests. Test number 1 is the acquired baseline used for assessing group homogeneity and the following numbers indicate performance test results after user training in each session. The red line indicates the progression of the test group and the blue line the progression of the control group.

The baseline performance test showed no difference between the two group, showing the two groups to be homogeneous at initiation. The Fitts' Law test results did not show any significant improvement over the three sessions for any of the five test measures for both the test and control group ($p > 0.05$). Similarly, there was no significant difference 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 covers the results acquired from measurements obtained during user training sessions. During user training subjects were instructed to train movements in being performed such that the control system recognized the movement as the actually performed movement. During this training the number of times subjects correctly performed an instructed movement to the contraction interval shown in the training interface was recorded, and will be referred to at the number of repetitions.

No significant difference in the total number of repetitions was found between sessions of either group ($p > 0.05$). When comparing the total number of repetitions of each session between groups accordingly, no significant difference were found either ($p > 0.05$).

An increased ability to reach the low intensities was found for the control group ($p < 0.05$, session 1 $= 16.13 \pm 5.59$, session 3 21.38 ± 6.78). Otherwise, similar results were yielded for both groups when comparing the subjects' ability to reach the three other contraction levels between sessions ($p > 0.05$).

No difference was found, when comparing the two groups' ability to reach different intensities during training either ($p > 0.05$). Comparing the ability to perform different movements during the training showed a significant improvement for the test group in ulnar deviation ($p < 0.05$, session 1 $= 11.38 \pm 4.27$, session 3 16.13 ± 2.95) and open hand ($p < 0.05$, session 1 $= 11.25 \pm 3.85$, session 3 $= 17.88 \pm 2.46$). A significant decrease in performance was found for the control group's ability to perform flexion ($p < 0.05$, session 2$= 16.63 \pm 2.77$, session 3$= 11.00 \pm 3.16$). Otherwise, no significant difference between the three sessions for the two groups was found ($p > 0.05$).

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 number of repetitions 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 ($p > 0.05$).

Cluster Dispersion and Separability Results

In this section results from the data acquisition are presented. The data used for training the LDA based classifier was examined. Each movement resulted in a cluster of data points, which was examined in order to analyse the change in cluster dispersion and distance between cluster centroids. For both groups the mean distance between the cluster centroids were calculated. The change in between cluster distances over the three sessions were tested and showed no significant difference ($p > 0.05$). Likewise, no significant difference in the development of cluster distances between the groups was found ($p > 0.05$).

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. The comparison between groups showed that the control group achieved a significant improvement of within cluster distances compared to the test group in session three ($p < 0.05$), where 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 objective of the study was to investigate if exposing subjects to user training, in which confidence levels of movement class recognition was used as feedback, would show statistically significant improvement in performance in a classification-based myoelectric prosthetic control scheme, when compared to subjects who received single-class feedback.

The results showed no significant difference between the test and control group within the Fitts' Law test, in all comparisons between and within groups ($p > 0.05$). This meant that no group performed better compared to the other, and that neither of the groups managed to improve significantly during the three sessions of training and testing. The only significant difference ($p < 0.05$) between the groups were found in the training when performing the closed hand motion, where the test group performed worse than the control group. This difference could be the result of the training type, the number of subjects or a faster learning ability within the control group.

A main cause of the lacking development within the groups could be the result of higher ID's ($4.88 - 6.41$) compared to other studies ($1.59 - 3.46$) [17, 22]. Several subjects struggled in reaching any targets, and if the subject was unable to reach any targets, all the Fitts' Law measures except CR were unusable in statistics. This lead 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 improvement of precision when completing the test.

At the same time a high ID led to subjects becoming frustrated when they had troubles reaching targets. When observing the test it was clear that this frustration resulted in the subjects forgetting how to perform precise movements, which then led to

further frustration. This factor could also have had an effect on the subjects performance. Significant improvement in development of movement precision might also take more than three sessions, and this could also be a cause of the lacking development of the subjects. In developing the understanding of precision there should also be a higher focus on rest, as this is a crucial part of the performance test. Some of the subjects did not understand the importance of returning to rest after a performed movement during user training, which might have been reflected in the performance test.
The above points should be taken into consideration in future studies when examining the use of confidence scores as visual feedback in user training to improve performance.

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 of the third session, 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 lead to the assessment that the control group became better at performing the exact movements during data acquisition when compared to the test group.

Optimization of Study

When implementing the performance test interface the ID's should be adjusted in order for the subjects to reach a CR of 80% to 100%, as reported in previous studies [17, 22]. This might yield a more clear indication of 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 carrying out the performance test. A problem observed during the Fitts' Law test was that subjects were affected by the cursor being reset to origo. The subjects' current movement were carried over during the transitioning between targets. A suggestion to future studies is to include a transition break of 1 second when a new target appears.
During user training the subjects should be forced to get back to rest, in order to train the ability to dwell within a target in the performance test. This requirement was not implemented in the current training interface, but the importance of learning to rest when using classifiers should be examined in future studies.
In future testing, the number of sessions should be more than three. In that relation it would be beneficial to examine the time it takes to improve performance in order to find the minimum number of sessions necessary to achieve higher precision when performing specific hand gestures.
At last a larger number of 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 session, while others struggled with reaching just one target during each session.

CONCLUSION

Based on the results in the experiment it was found that training the user with confidence score feedback compared to single-class feedback can not be linked to any significant difference improvement in performance evaluated through a Fitts' Law test. Furthermore, no significant improvement during a three day training period for either the control or the test group was detected.

Contrarily, it appears that training the user with single-class feedback can lead to a closer clustering of EMG data compared to training with confidence score feedback. This can be concluded, as a significant improvement was found between the first and last dataset recorded for the control group. To further support this the EMG-signal of the subjects who received single-class feedback clustered significantly closer 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 for providing constructive feedback, and the School of Medicine and Health at Aalborg University for providing equipment and the facilities to complete this study. Additionally, the authors are very thankful for all the voluntary participants.

References

[1] Bart Peerdeman et al. "Myoelectric forearm prostheses: State of the art from a user-centered perspective". In: *The Journal of Rehabilitation Research and Development* 48.6 (2011), p. 719.

[2] 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.

[3] 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.

[4] 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.

[5] 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.

[6] COAPT. *Coapt engineering*. 2018.

[7] I Mendez et al. "Evaluation of the Myo Armband for the Classification of hand motions". In: *International Conference on Rehabilitation Robotics* (2017), pp. 1211–1214.

[8] 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).

[9] Simon Bruun et al. "The effect of limb position on myoelectric prosthetic control using linear regression". In: *AAU* (2017).

[10] 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.

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

[12] 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).

[13] 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.

[14] Yinfeng Fang et al. "Classifier-Feedback-Based User Training". In: *IEEE Transactions on Biomedical Engineering* 64.11 (2017), pp. 2575–2583.

[15] 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).

[16] Ann M. Simon, Blair A. Lock, and Kathy A. Stubblefield. "Patient training for functional use of pattern recognition controlled prostheses". In: *Journal of Prosthetics and Orthotics* 24.2 (2013), pp. 56–64.

[17] 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.

[18] Fernando D Farfán, Julio C Politti, and Carmelo J Felice. "Evaluation of EMG processing techniques using Information Theory". In: *BioMedical Engineering OnLine* 9.1 (2010), p. 72.

[19] 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.

[20] B. Hudgins, P. Parker, and R.N. Scott. "A new strategy for multifunction myoelectri control". In: *IEEE Transactions on Biomedical Engineering* 40.1 (1993), pp. 82–94.

[21] Angkoon Phinyomark, Pornchai Phukpattaranont, and Chusak Limsakul. "Feature reduction and selection for EMG signal classification". In: *Expert Systems with Applications* 39 (2012).

[22] 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.

[23] 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.