Examining if Confidence Score Feedback During User Training Can Improve Users' Ability in Controlling Upper Limb Prosthetics

Simon Bruun*, Oliver T. Damsgaard*, Martin A. Garenfeld* and Christian K. Mortensen*

** Undergrad. Student, AAU*

Abstract

The user rejection rate of myoelectric prosthetics is currently high, due to slow and inaccurate control. 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 control 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. 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. The results from the Fitts' Law test showed no significant difference between the two groups and no improvement over the three sessions for either of the groups. Overall, three sessions of user training with confidence score feedback showed to be an insufficient training period to observe a significant improvement within and between subject groups.

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

I. INTRODUCTION

The loss of any part of a limb or limb as a whole is a debilitating condition 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 an upper limb prosthesis. [**Peerdeman2011**]

In recent years, myoelectric prosthetics have become increasingly better in performance in a clinical environment, where the number of DOF's in prosthetic hands have increased. However, the ability to control these DOF's are limited by the need for more complex control systems, leading to a lack of functionality in when used in daily life tasks. In addition to the restricted control, the discomfort of prosthetics are causing patients to reject their provided prosthetic device, with a 23-45 % rejection rate. [**Biddiss2007**] Commercially available prosthetics range from passive cosmetic prosthetics to functional few degrees of freedom (DOFs) cable-driven prosthetics and more advanced myoelectric controlled prosthetics.

In recent years several complex multi DOF prosthetic hands have been developed. Examples of these 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. [**Belter2013**] Despite the efforts to advance and improve the functionality of prosthetics, a critical bottleneck still exist: the ability to properly control the prosthesis [**Hwang2017**].

Most commercially available myoelectric controlled prosthetics rely on manually switching between different DOFs in the prosthetic [**Fougner2012**]. This is a robust control scheme, but it is a slow and non-intuitive in movement. In the research area of myoelectric prosthetics, newer control schemes have been developed. These control schemes are classification- or regression-based. Classification have been used for many years in research, but is to date only used in one commercially available prosthesis [**Coapt2018**]. 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 [**Fougner2012**, **Scheme2015**]. The regression control scheme determines the output signal for an input based on a regression model. This provides a continuous output value, facilitating simultaneous control contrary to classification which provides only a single class output at a time. [**Hwang2017**, **Hahne2014**] In both control schemes, 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. [**Powell2014**]

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 [**Fougner2012**]. This area focuses on the design of the hardware and software side of the system in EMG prosthetics. The awareness in the research area shows a very single-minded approach to possible improvements of control, and thus mainly system training have been researched [**Jiang2012**]. Jiang et al. [**Jiang2012**] 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 [**Powell2014**, **Fang2017**, **Pan2017**]. Contrary to system training, user training focuses on the user's ability to control a prosthesis [**Fougner2012**]. 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. This is carried out in order to train the user in performing more distinguishable movements, which facilitate better control. 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. [**Powell2014**, **Simon2013**]
In a 2014 study, Powell et al. [**Powell2014**] provided amputee users with real-time visual feedback of an animated prosthesis. This feedback provided the user with visual information on how the prosthesis would move related to which contractions the user performed. Through a 10 session experiment the subject group experienced an average movement completion percentage from 70.8 % to 99.0 %. Fang et al. [**Fang2017**] provided real-time visual feedback of subjects' performed movement in relation to the classes defined in the system. The feedback visualized a 2D map of cluster centroids based on a PCA of the EMG. The users were instructed in matching the centroid of an online PCA based cursor to the centroid on the 2D map corresponding to the performed movement. When subjects could match the cursor to the centroid of a cluster the performed movement corresponded the best with the class of that movement. The study demonstrated steady improvement of hand motion accuracy compared to conventional labelled feedback. [**Fang2017**] Based on the findings of these studies other training schemes should be examined in order to find whether other methods could improve user performance further.

A 2013 study by Scheme et al. [**Scheme2013**] 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. [**Scheme2013**] 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 by showing them which classes the system recognizes. This would give the possibility for the user to modulate their EMG patterns in order to increase the confidence of the classifier.

Based on the presented possibilities in improving user training, this study will seek to investigate the use of confidence scores as 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 are provided as the feedback, will show statistically significant improvement in performance in a classification-based myoelectric prosthetic control scheme, when comparing to subjects receiving discrete class label feedback [**Fang2017**].

The remainder of this paper will cover the following; Section II will cover the calculation of confidence scores. Section III will cover how the study design was structured in order to test the application of confidence scores in user training. Afterwards the three interfaces for data acquisition, user training and performance test will be presented respectively. Section IV will present the findings and in section V the findings will be discussed and analysed. Section VI concludes the study.

II. BACKGROUND

As the application of confidence scores during user training is the main focus of this study, a brief derivation of confidence scores will be presented in this section. This theoretical derivation of confidence scores from a LDA classifier is based on a study by Scheme et al. [**Scheme2013**].
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 [**Scheme2013a**], 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 that 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 Σ. Σ_j can thus be replaced with the common covariance matrix Σ. 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:

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

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. The LDA classifier will be used in the control scheme. To obtain smoother control, the class with the highest average likelihood based on features from the previous three segments is chosen as output class.

III. 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 to 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. All implementations have been performed using MATLAB (2017b).
The difference between the test and control groups, and the main area of interest in the study, was in the feedback provided during user training. The test group received the estimated probabilities of each class (confidence scores), while the control group only received label feedback (the estimated class). A flowchart of the study design can be seen in figure 1.

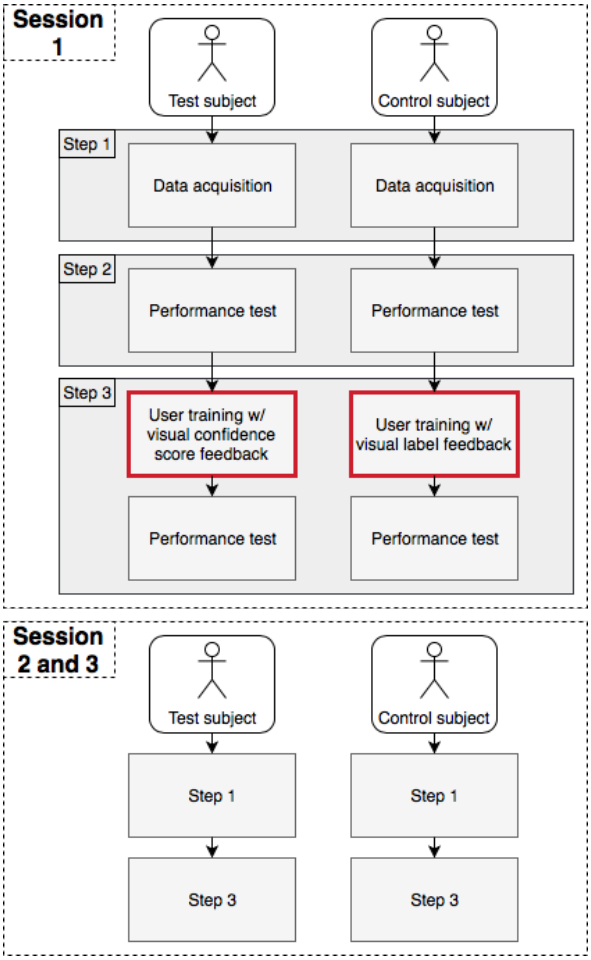

Fig. 1: Graphical illustration of the experiment showing the steps of each session for the test and control group. Highlighted is user training in step 3 which was the only step that varied between the two groups, and comprised the main area of research interest in the experiment.

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 recordings that can be classified with significantly similar accuracy as EMG recordings acquired with conventional EMG surface electrodes sampled at 1000 Hz [**Mendez2017**].

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 had their dominant forearm disinfected, and were instructed in wearing the MYB at the thickest part of that forearm. 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.

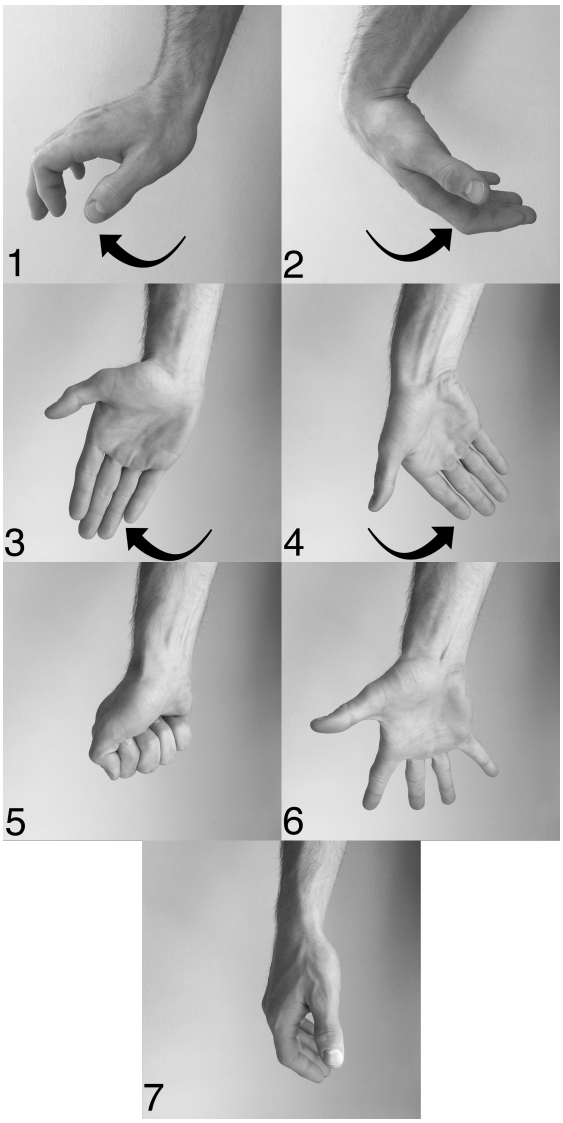

Fig. 2: Illustration of the movements performed in the experiment. 1: Wrist extension, 2: Wrist flexion, 3: Radial deviation, 4: Ulnar deviation, 5: Closed hand, 6: Opened hand, 7: rest.

According to Scheme et al. [**Scheme2015**], 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.

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. [**Farfan2010**]. Based on using the MYB for data acquisition recommendations made by Donovan et al. [**Donovan2017**] 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 [**Hudgins1993**] in a LDA based control scheme [**Donovan2017**].
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 a portion of the features Donavan et al. [**Donovan2017**] 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 [**Phiny2012**].

Proportional Control

Classification outputs the desired movement but not the intensity of that movement. Therefore, to estimate the intensity, 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 signal generated when tracking the trapezoidal trajectory during the data acquisition. Thus, the proportional output value was a single value between 0 and 1. The calculation was as follows:

$$\hat{Y} = \alpha + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_i X_i + \varepsilon_i \quad (7)$$

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, β is the regression slope and 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 was being recognized by the control system. The only difference between the test and control group was the feedback received in the vertical bar plot. The test group received confidence score feedback (multiple bars) as seen in figure 3 and the control group received label feedback (single bar) as seen in figure 4.

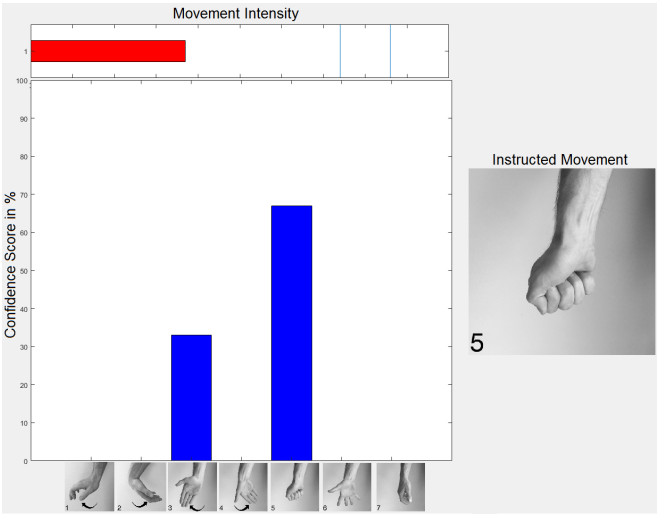

Fig. 3: Illustration of the user training interface for the test group. The vertical bar plot indicates which movement is being recognized visualized by the images of each movement; a full bar corresponds to 100 % recognition confidence. The horizontal bar plot indicates contraction level, where a full 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 test group received confidence score feedback by having the possibility of multiple bars being shown.

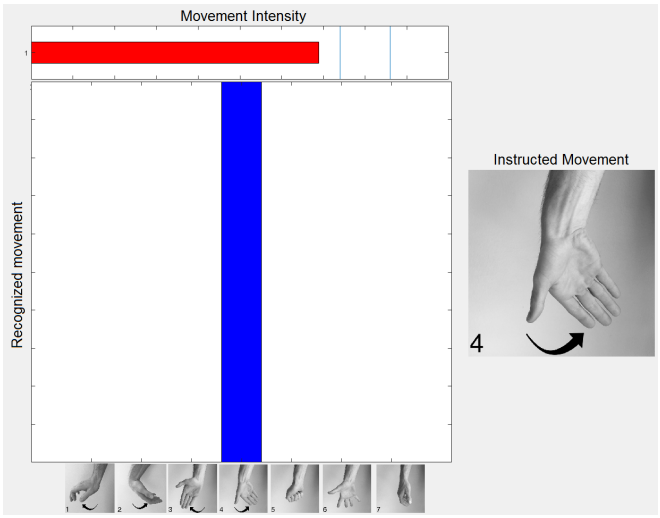

Fig. 4: Illustration of the user training interface for the control group. The same interface used for the test group was presented to the control group, except the feedback instead consisted of label feedback. Thereby the control group were only presented with the most certain recognized movement, shown with a single bar.

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. Basically, the training should motivate the subjects to modulate their contractions to maximize the classifier confidence. The subjects should perform contractions such that the probability bar for the target class is maximized while the probability bars for the other classes are minimized. 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 for the test group and the correct class for the control group, 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. A task completion was referred to as a repetition and was saved as a user training outcome measure. 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 [**Scheme2013**, **Scheme2013a**]. The user controlled a circular cursor in a Cartesian coordinate system, where the cursor was to be matched with the appearing targets. Extension/flexion of the wrist moved the cursor horizontally, radial/ulnar deviation moved the cursor vertically and opened/closed hand increased/decreased the size of the cursor. The cursor moved proportional to contraction intensity with a velocity between 0 and 1, where 1 corresponded to the MVC. An illustration of the Fitts' Law test interface can be see in figure 5. 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 origin. 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.

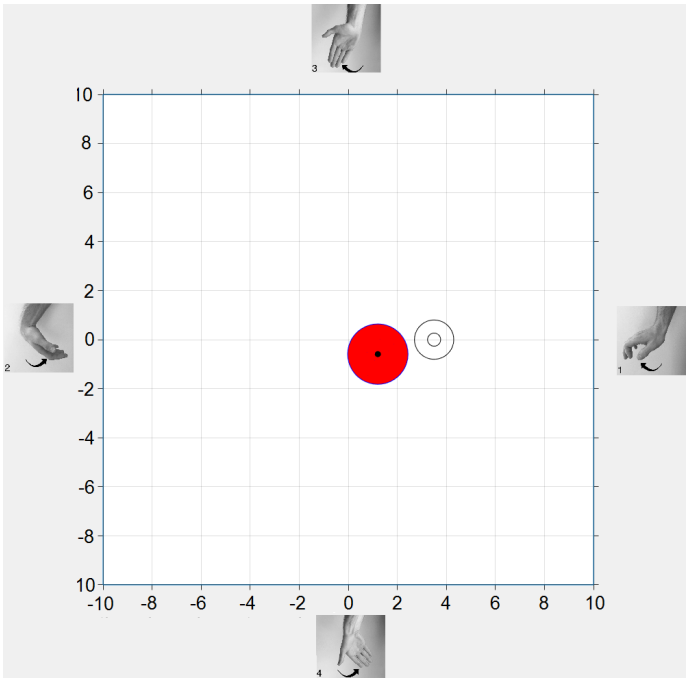

Fig. 5: The implemented interface for the modified Fitts' Law test. The user controlled the red cursor with the centred bold mark. The target consisted of a circle with a larger circle surrounding it. The user was instructed in matching the cursor with the target, where the bold mark should be positioned 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, and blue when the dwell time was reached.

Originally the Fitts' Law test had a single performance measure, *throughput* (TP) [**Fitts1954**]. 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:

$$TP = \frac{1}{N}\sum_{i=1}^{N}\frac{ID_i}{MT_i} \tag{8}$$

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

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

According to [**Scheme2013a**], 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 [**Scheme2013**, **Scheme2013a**]. 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.

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

$$C = \frac{\sum_{n=1}^{N} a_n, b_n, \ \ldots \ k_n}{N} \tag{10}$$

Where C is the centroid, n is the number of data points 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:

$$ED(p,q) = \sqrt{(p_1-q_1)^2+(p_2-q_2)^2 \ + \ \ldots \ +(p_k-q_k)^2} \tag{11}$$

Where p_k and q_k are the coordinates of vectors p and q respectively. 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.

Statistical Analysis

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.

IV. 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. The plotted mean and standard deviations of each measure for all subjects in each group in the performance test for each session can be seen in figure 6.

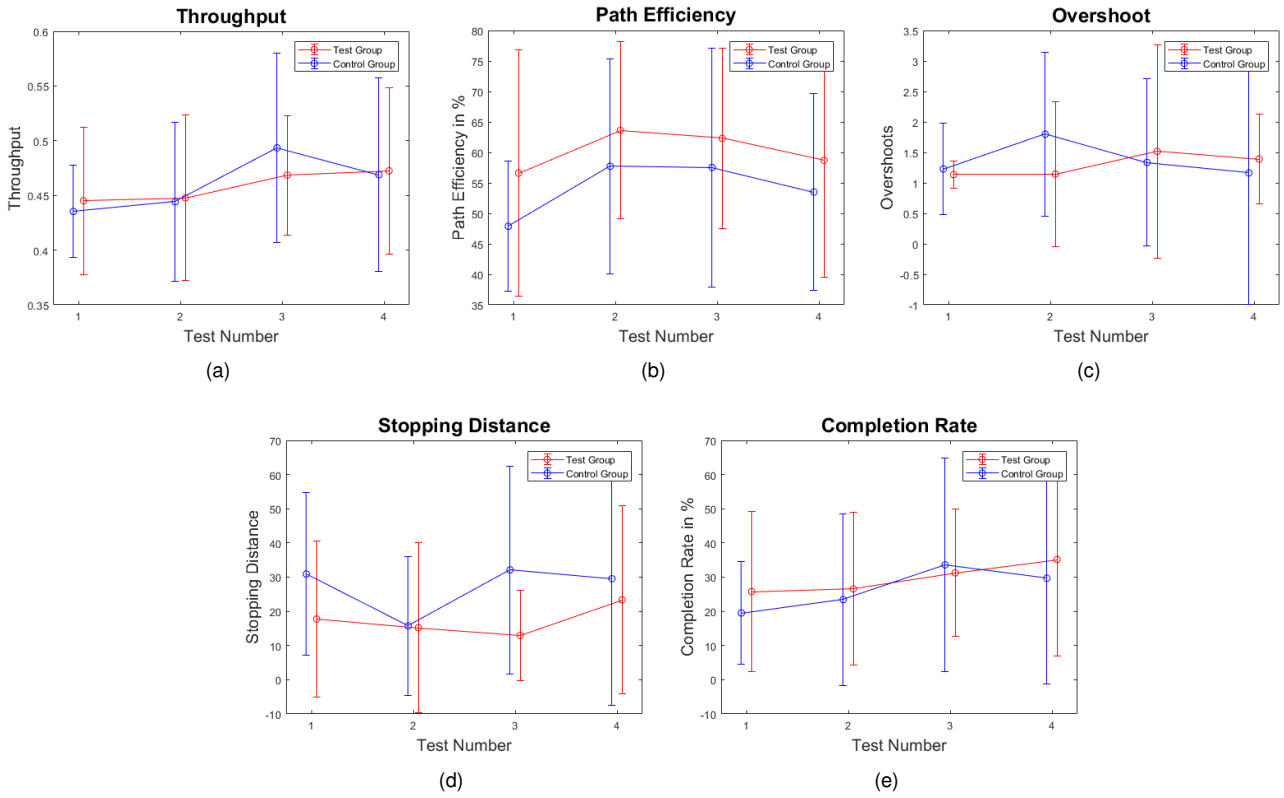

Fig. 6: Figure illustrating the five performance measures; *a) Throughput, b) Path efficiency, c) Overshoot, d) Stopping distance, e) Completion rate*, used for quantifying user performance across all four tests. Test number 1 is the acquired baseline used for assessing group homogeneity and the following 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 outcome measure results 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.

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 get repetitions in the low intensities was found for the control group ($p < 0.05$, session 1 $= 16.13 \pm 5.59$, session 3 $= 21.38 \pm 6.78$). 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 $= 16.13 \pm 2.95$) 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 arb. unit, and session three was 323.43 ± 171.13 arb. unit. 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 arb. unit, while the control group had 323.43 ± 171.13 arb. unit.

V. 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 label 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 user 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. The control group improved in number of repetitions in lower intensities during user training between session 1 and 3. It was expected that this improvement would be found within the test group, as the low intensity motions where the motions from which the classifier would get most confused, and the confidence score then would provide insight in how to correct the movement best possible. A reason for the contradictory result might be that the subject were confused by the confidence scores and found the information excessive.

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 ($1.59 - 3.46$) [**Scheme2013**, **Scheme2013a**]. 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 resulted in a weaker statistical test, as fewer observations were included in the comparison of the remaining measures. In addition, 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 de-

velopment 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 ($p < 0.05$).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 and minimum number of DOF's used to reach a target, should be lowered in order for the subjects to reach a CR of 80% to 100%, as reported in previous studies [**Scheme2013**, **Scheme2013a**]. 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 origin. 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. This was also found in [**Pan2017**], who similarly did not achieve significant improvement in performance following a short time user training intervention, whereas Powell et al. [**Powell2014**] found a steady improvement during a 9 session user training study. 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.

VI. CONCLUSION

Based on the results in the experiment it was found that training the user with confidence score feedback compared to label 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. These findings are most likely due to the high index of difficulty, making it hard to draw any conclusions based on the Fitts' Law test.

Contrarily, it appears that training the user with label 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 label feedback clustered significantly closer than the test group on the last day of testing. This shows that training based on confidence scores might not be a way to improve performance, which should be examined further by the use of Fitts' Law tests with lower ID's and a higher number of training sessions.

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.

VII. APPENDIX

Features In this section the equations used for calculating the features used in this project.

MAV is a feature that primarily is affected by the force produces when making a contraction. MAV is extracted for each

window and calculated for each of the i^{th} channel. The extraction is expressed as:

$$MAV_i = \frac{\sum_{n=1}^{ws} |x_i[n]|}{ws} \tag{12}$$

where ws is the window size, the number of raw data points in that exact window. $x_i[n]$ is the n^{th} raw data points from the i^{th} channel.

The mean MAV across all channels, MMAV, is used to remove dependency of movement intensity. MMAV is calculated by using the MAV of all channels for the current window, and is done as following:

$$MMAV = \frac{\sum_{i=1}^{8} MAV_i}{8} \tag{13}$$

MMAV can be used to scale the MAV feature creating the SMAV feature. This feature should represent a non-dimensional relationship between channels. SMAV is simply calculated as:

$$SMAV_i = \frac{MAV_i}{MMAV} \tag{14}$$

As each of the eight EMG sensors in the MYB are located around the arm, they acquire signals from a mixture of sources. Also individual sources may affect multiple sensors depending on their size. Due to this a source measured by multiple sensors will effect their acquired signal correlation. An idea is therefore to calculate the correlation coefficient between each channel and its neighboring channel.

$$CC_i = \frac{\sum_{n=1}^{ws} X_i[n] X_{i+1}[n]}{ws} \tag{15}$$

$X_i[n]$ is the n^{th} normalized data point from channel i. When calculating CC the data from each window is normalized by subtracting its mean value from each raw data point, and afterwards divided by their standard deviation.

Calculating CC can prove rather demanding in computational power due to the series og multiplication operations. Therefore Donovan et al. [**Donovan2017**] proposed introducing a mean absolute difference-based feature of lower computational complexity which still characterizes the spatial relationship between channels. The MAD feature is normalized in the same way as CC, making up the MADN feature calculated as:

$$MADN_i = \frac{\sum_{n=1}^{ws} |X_i[n] - X_{i+1}[n]|}{ws} \tag{16}$$

If the normalization of the signal proves too demanding the feature can be calculated on the raw EMG-signal without the normalization. This makes up the MADR feature, calculated as:

$$MADR_i = \frac{\sum_{n=1}^{ws} |x_i[n] - x_{i+1}[n]|}{ws} \tag{17}$$

As the SMAV feature the MAD feature can be scaled by MMAV to remove movement intensity dependency. SMADR is calculated for each channel by:

$$SMADR_i = \frac{MADR_i}{MMAV} \tag{18}$$

As stated in the beginning some of these features introduce redundancy, subsequently the features of SMAV, CC, MADN and SMADR are the ones used for classification. [**Donovan2017**]

To further improve the decision foundation of the classifier it was proposed to include the time domain feature of WL calculated by:

$$WL_i = \sum_{n=1}^{N-1} |x_{i+1}[n] - x_i[n]| \tag{19}$$

WL is a measure of the signal complexity by calculating the cumulative length for each channel [**Phiny2012**].

Fitts' Law Measures In this section the equations for the Fitts' Law measures are presented. Throughput (TP) which represents the trade-off between speed and accuracy. TP uses the relationship of time taken to reach a certain target in seconds (MT) and the index of difficulty (ID). This forms: [**Scheme2013**, **Fitts1954**]

$$TP = \frac{1}{N} \sum_{i=1}^{N} \frac{ID_i}{MT_i} \tag{20}$$

where i is a specific movement and N is the total number of movements. ID relates to the target distance D and width W. The ID for each task, from the origin to a specific target of a certain size is calculated using [**Scheme2013**, **Fitts1954**]:

$$ID = log_2(\frac{D}{W} + 1) \tag{21}$$

Path Efficiency (PE) describes the quality of control by making a measure of the straightness of the cursor's path to the target, by making a ratio of the actual path distance versus the optimal path distance. This tests the users ability to continuously control the cursor position. Following the optimal path will result in a PE of 100%. PE is calculated as follows [**Scheme2013**, **Poulton2013**]:

$$PE = \frac{Optimal\ Distance}{Actual\ Distance} \tag{22}$$

Overshoot (OS) is the number of times the cursor enters and then leaves the target before the dwell time inside the target is reached, across all target in the task, divided by the total number of targets. OS tests the users ability to control the velocity of the cursor accurately. A perfect OS-score of zero is reached if the cursor dwells within the target boundaries on the first try for all targets, and is calculated as the following [**Scheme2013**, **Poulton2013**]:

$$OS = \frac{Total\ Number\ of\ Overshoots}{Total\ Number\ of\ Targets} \tag{23}$$

Stopping Distance (SD) describes the users ability to rest and thereby perform no movement. The SD measure is the distance moved during the dwell time across all targets, and is given as [**Scheme2013**]:

$$SD = \sum_{i=1}^{N} (Distance\ Inside\ Target)_i \tag{24}$$

where i is a reached target and N is the total number of reached targets.

Completion Rate (CR) describes the percentage of targets reached within the total allowed time. This gives a general idea of the user's performance, and is calculated as [**Scheme2013**, **Simon2011**]:

$$CR = \frac{Number\ of\ Reached\ Targets}{Total\ Number\ of\ Targets} \tag{25}$$