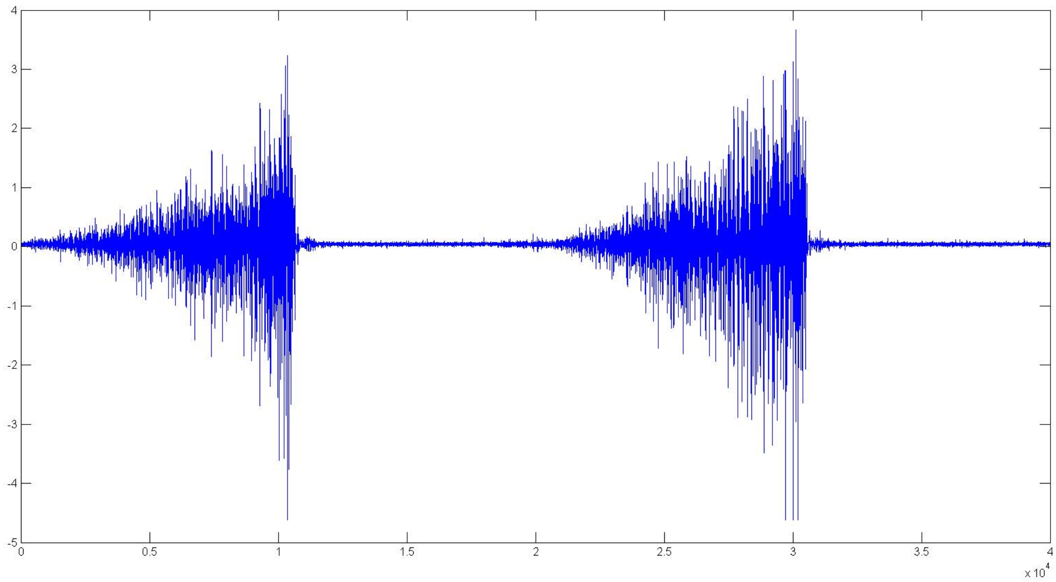
CHALMERS



Effect of bioelectrical signal acquisition on classification performance

Master of Science Thesis

PONTUS LÖVINGER

Department of Signals and Systems
Division of Biomedical Engineering
CHALMERS UNIVERSITY OF TECHNOLOGY
Göteborg, Sweden 2014
Report No. EX014/2014

REPORT NO. EX014/2014

Effect of bioelectrical signal acquisition on classification performance

PONTUS LÖVINGER

Department of Signals and Systems
Division of Biomedical Engineering
CHALMERS UNIVERSITY OF TECHNOLOGY
Göteborg, Sweden 2014

Effect of bioelectrical signal acquisition on classification performance

PONTUS LÖVINGER

Technical Report no. EX014/2014
Department of Signals and Systems
Division of Biomedical Engineering
Chalmers University of Technology
SE–412 96 Göteborg
Sweden

Chalmers Bibliotek, Reproservice
Göteborg, Sweden 2014

Effect of bioelectrical signal acquisition on classification performance

PONTUS LÖVINGER
Department of Signals and Systems
Division of Biomedical Engineering
Chalmers University of Technology

Abstract

Myoelectric robotic prosthetic devices are assistive devices for people with amputations. These devices can be controlled by training a myoelectric pattern recognition (MPR) algorithm with myoelectric signals (MES) sampled from electrodes placed over muscles. Conventionally the MPR algorithms have been train with MES data produced when performing a movement with sustained contraction force. In this thesis, the effect of training the algorithms with the contraction force increasing as a ramp over time has been investigated. For this, the offline accuracies as well as real-time accuracies have been analyzed in order to evaluate the performance of classifiers trained with ramp or sustained contraction data. Two pattern recognition algorithms (LDA and MLP) have been compared for real-time classification. From the offline accuracy it was found that the average accuracy is higher when training classifiers with sustained contraction data compared to ramp contraction data. The average accuracy for real-time classification was found to be similar (no statistically significant difference) when comparing classifiers trained with sustained and ramp contraction data. Of the two algorithms compared for real-time classification using ramp data, linear discriminant analysis (LDA) was found to be better than multilayer perceptron (MLP). From the data it was not possible to find any connection between the offline and real-time accuracies except than that the offline accuracies are high than the real-time accuracies.

Index Terms: Myoelectric prosthesis, classifier, dynamic training, ramp training, sustained contraction, myoelectric pattern recognition

Acknowledgements

This work has been carried out at the Department of Signals and Systems at Chalmers University of Technology. Integrum AB has supported me with a workplace and lab equipment to perform the necessary experiments and studies. I would like to thank my advisor Max Ortiz Catalan for helping me and guiding me through the work process. I would also like to express my gratitude to the people that devoted their time to help me with the experiments conducted.

Pontus Lövinger
Göteborg, Sweden, 2014

List of abbreviations

MPR = Myoelectric pattern recognition
tmabs = Mean Absolute
twl = Wavelength (time domain)
tzc = Zero Crossing
tslpch = Slope Changes
fwl = Wavelength (frequency domain)
trms = Root mean square
tstd = Standard deviation
fmd = Median frequency
MLP = Multilayer Perceptron
LDA = Linear Discriminant Analysis
RFN = Regulatory Feedback Networks
f1 = {tmabs, twl, tzc, tslpch}
f2 = {tmabs, fwl, trms, tzc}
f3 = {tstd, fwl, fmd}

Contents

Chapter 1

Introduction

Most people want to be healthy and be able to do everyday things such as pouring themselves a glass of water or a cup of coffee. The greater part of the population takes these things for granted, but there is a certain group that lack the possibility to perform these everyday tasks. People with disabilities, like upper limb amputees, cannot use some devices which require intact limbs. Myoelectrically controlled robotic prosthetic hands or arms are assistive devices which purpose are to restore some of the function which an intact arm provides and thus aid the amputee in different situations.

Even though advancements has been made in the field of myoelectrically controlled upper limb robotic prosthetic devices, there is still a wide gap between what is currently the cutting edge technology and what would be needed in order to be beneficial for the greater part of the amputee population [1]. Since the first clinically available myoelectric devices were presented back in the 1960's the area of upper limb myoelectric prostheses has not been subjected to any revolutionary advancements, but the advancements have rather been small steps forward [2]. An investigation of literature showed that the characteristics that the amputees found most problematic about the commercially available prosthetic devices were: "non-intrusive control, lack of sufficient feedback, and insufficient functionality" [3]. These problems reflects the need to develop devices that are more usable and can better suit the need from amputees in order to increase their quality of life.

Chapter 2

Background

The earliest myoelectric prosthetic devices emerged in the 1950 and 1960 [1]. These could for example be controlled by using a threshold which the myoelectric signals (MES) has to exceed in order to active a certain movement. The systems were very limited and would not allow the user to perform more than a few movements. Even though these old systems are so limited in what they can do, this is still the way most of the commercially available devices work today [1].

Our brain produces nerve signals to control the contraction of our muscles [2]. Amputees also generate these signals that are meant to reach muscles in the body even though a person is missing a part of a limb. If an amputee still has parts of the muscles related to a specific movement, these muscles will contract if the person tries to perform such movement. However, if a person has not used the targeted muscles in a long time he or she might need some time to get used to using those muscles again. As the muscles contract MES are generated and by using surface electrodes, the MES can be recorded from the contraction of the muscles and be used to train and control a robotic prosthetic device. Even If the appropriate muscles for a certain movement are missing, the prosthesis can still be controlled by using other muscles than those normally related to a certain motion [4].

Since the first prosthetic devices were developed, more advanced systems have been presented. One example is the systems that rely on myoelectric pattern recognition (MPR). When using electrodes the signal recorded from one electrode is not only the signal from the muscle directly under the electrode. Because of cross talk between muscles, the MES recorded generally include signals from several muscles [4]. Because of this, controlling a robotic prosthesis is not as easy as just to put an electrode over a muscle and relate it to some movement, at least not if several movements should be used. MPR based systems tries to recognize what the intended movement is by recording the MES from different muscle locations and characterizing the signals into different features that the signal has. Before a MPR based prosthesis can be used for prediction, a classifier has to be trained in order to recognize the movements that the classifier should be able to distinguish between. This is done in the following way:

1. MES are recorded and preprocessing of the recorded data is performed. Preprocessing can include for example filtering to remove power line harmonics.

2. Next the data is segmented by windowing, which divides the signal into time windows. Time window lengths of 32 ms [5] up to 256 ms [6] have been used or investigated in different papers. The length of the time windows is a tradeoff between prediction accuracy and responsiveness. Since a longer time window contributes with more information, the prediction of each movement is more accurate or more likely to be the correct one, but at the same time the system response is slower since it takes longer to receive and process the necessary information. When the time windows are divided, it is common that each window has an overlap so that the start of each windows is the same as the last part of the previous window.

3. Feature are then extracted from the time windows. The features extracted from the MES fall into one of three categories which are time domain, frequency domain or time-frequency domain features [2]. Time domain features can for example include mean absolute value, root mean square or zero crossing [3] [4]. Frequency domain feature can include power spectrum properties [7] and time-frequency domain feature can be for example wavelet transform or wavelet packet transform [2].

4. When the signal features have been extracted from the windows, the features are used to train a classifier to be able to distinguish between the intended movements. Different type of algorithms have been tested by researchers and some examples of algorithms that have been used are those that rely upon artificial neural networks (ANN) [8], support vector machines [9], Gaussian mixture models [10], hidden Markov models [11], fuzzy c-means [12] and linear discriminant analysis (LDA) [6].

Sometimes dimensionality reduction is used before the feature vector is fed into the classifier. This operation reduces the dimension of the feature vector and tries to preserve the most informative parts of the features. In this way a smaller feature vector can be fed into the classifier, which reduce the computational load, but still preserve the most informative information of the original data [13].

When a classifier has been trained it can be used to predict movement in offline mode or in real-time. For offline mode the classifier is fed with already recorded data that is divided into time widows and where the feature are extracted from the data. When features have been extracted and are fed to the classifier, it uses the information to figure out which movement the data most likely represents. When performing real-time pattern recognition the data is continuously recorded. When this is done predictions will be made with a certain time delay depending on the length of the time window. In order to fill a time window the system has to wait until enough data has been sampled to fill a window and then feature from that window can be extracted and fed into the classifier for prediction.

There exists a relation between the amplitude of the MES and the intended force of a certain movement [14]. Signals recorded from the stump of the amputee can be utilized for proportional control of an artificial limb by scaling the force or speed of the prosthesis with the contraction force. The proportional control is most often implemented using the average effort, over all the channels used for recording of the MES, to scale signal input to the actuators of the prosthesis [1].

As mentioned in the introduction, amputees find the lack of feedback problematic. It can be demanding for the patients to control a myoelectric prosthetic device since the amputees cannot feel, for example, how hard they are closing their hand or how fast they are performing a certain movement. The only feedback they have is the visual feedback from the prosthesis. Different solutions for implementing feedback exist, but the main focus of research has been on vibrotactile and electrotactile stimulation [3]. Implementing for example force feedback could increase the intuitiveness of controlling a prosthesis when performing grasping motions and decrease the attention needed when performing these motions [3]. Force feedback can also provide the user with information about the stiffness of an object and make handling of fragile object easier. Feedback of limb positioning has also been tested by using stimulation applied to the skin that is proportional to the position of the elbow. With this feedback the user showed increased performance when performing reaching and matching tasks [3].

The data used when training a classifier is usually recorded from movements performed with a sustained contraction force. For example the subject may be asked to perform ”open hand” for three seconds, without any specific instruction on which force the movement must be performed with. If the subject during real-time control would perform a movement not with the same force as when performing the movement in the training phase, the risk increases that the movements will be misclassified as some other movement, especially since many classifiers are trained with features that depend on the signal amplitude [15]. In this thesis the effects of training classifiers with data which amplitude increases linearly as a ramp over the recording has been investigated. In order to obtain more accurate predictions for movements performed with different contraction forces, the classifiers can be trained with data recorded from several effort levels. In order to obtain data from many different effort levels the recordings are perform with the effort increasing linearly over time. If each movement is recorded as a ramp the recordings should contain more information from many different effort levels, and thereby also allow the classifier to generate more accurate predictions for different effort levels. This could be advantageous from the practical point of view since it might be easier to get more correct prediction when preforming the different movements with varying force, like we do in our everyday life. The idea that training the classifier with several effort levels should generate better accuracy when testing with several effort levels has been studied by Scheme *et al.* [15] and Al-Timemy *et al.* [16].

2.1 Previous work

In the work by Scheme *et al.* [15] able bodied subjects were used to evaluate the offline and real-time classification error for different cases.

- LDA classifier with feature set {mean absolute value, zero crossing, slope sign changes, waveform length} was used for all experiments.
- Offline classification error:
  - Training classifiers with sustained contraction data with different forces (low, medium, high, 20%, 30%, 40%, 50%, 60%, 70% and 80%) and testing against every force against itself and every other force (all combination).
  - Training with sustained contraction data form all forces and testing against ramp contraction data.
  - Training classifiers with ramp data and testing against all sustained contraction forces and against ramp data.

The results suggests that the classifier trained with ramp data had lower average classification error than the classifiers trained with steady contraction data.

In the study performed by Al-Timemy *et al.* [16] two amputees participated to evaluate the offline classification error for different cases.

- LDA classifier with feature sets {integral absolute value, waveform length, zero crossing, slope sign change, kurtosis} and {4th order auto regression, mean square}
- Offline classification error:
  - Training with sustained contraction data with low, medium and high contraction force and testing against all contraction forces (all combination).
  - Training with sustained contraction data with low, medium and high contraction force at the same time and testing against low, medium and high contraction force.

The results suggested that training a classifier with all force levels at the same time reduced the average prediction error compared to when training classifiers with only one contraction force and testing with other levels.

Chapter 3

Focus of the thesis

The previous studies has shown that training classifiers with data from several contraction forces has resulted in lower average prediction error when testing with data from different effort levels. The study performed by Scheme *et al.* also showed that the real-time classification error was improved when training with several effort levels by using ramp data. The effect of algorithm choice (LDA, MLP or RFN) on offline accuracy when training a classifier with ramp data has been investigated in this thesis. The relation between offline and real-time accuracy has been investigated. The real-time accuracy for an LDA classifier trained with steady contraction data has been compared with an LDA and MLP classifier trained with ramp data, in order to investigate if training with ramp data has an effect of the overall real-time accuracy. For this, ten different movements have been used and the offline accuracy for different feature sets, topologies and classifiers has been investigated.

3.1 Limitations

No effort has been put into learning how the pattern recognition algorithms work. Since all the algorithms used are already implemented in BioPatRec, this tool has been used.

The average accuracy for different set of movements have been investigated (10, 6 and 4 movements). However, no classifiers have been trained specifically for recognizing only the movements in the subset of movements (6 or 4 movements). The set of movements are introduced in the method section.

Chapter 4

Method

4.1 Subjects

Six able bodied subjects (two male and four female) participated in the experiments conducted and presented in this thesis.

4.2 Electrode setup

Disposable Ag/AgCl electrodes with diameter of 1 cm were used. The electrodes were placed in a bipolar configuration (two electrodes per channel) with two centimeter separation between each electrode pair. The MES were recorded from the most proximal third of the forearm using four channels (eight electrodes). Four channels were chosen because this has proven to be sufficient [6] [13] [17]. The first channel was placed along the extensor carpi ulnaris and the rest were equally spread around the forearm, see figure 4.1.

4.3 Equipment and software

The biopotential amplifiers used to amplify the MES are differential amplifiers with a gain of 2500. They are also embedded with active filters: 4th order high pass filter with cut off frequency at 20 Hz and a 2nd order low pass filter at 400 Hz. A notch filter at 50 Hz is also included to reduce the interference from power line harmonics. National Instruments ”NI USB-6212” DAQ with a resolution of 16 bits and sampling frequency at 2000 samples per second was used for sampling of the amplified MES signals. A platform for recording and processing myoelectric data call BioPatRec [18] was used. BioPatRec is a graphical user interface (GUI) based program written in MATLAB. It is a research platform containing source code to handle everything from signals acquisition to real-time control of prosthetic devices.

4.4 Recording sessions

Sustained contraction force. The standard recording session was performed by instructing the subjects to perform 10 movements (open hand, close hand, flex hand, extend hand, pronation, supination, side grip, fine grip, agree and pointer) in the order just mentioned, see figure 4.2. Every movement was performed with a sustained contraction force two times in series. The subjects were instructed to perform the movements with a normal contraction force that was subjectively interpreted by the subjects. Every movements was performed for 5 seconds with a 5 seconds resting period between the repetitions and with 8 seconds rest between new movements.

Ramp contraction. The ramp recording session was used to record signals that increase as a ramp throughout the recording. It started by obtaining the minimum voluntary contraction force by instructing the user to relax. Next, the maximum voluntary contraction (MVC) force for each movement was requested. The subjects were asked to execute the same movements as for the standard recording session, in the same order,

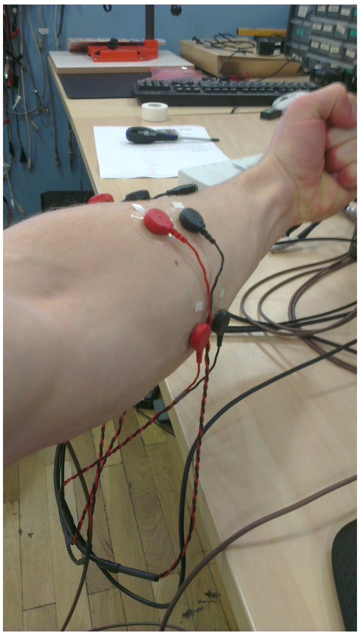


Fig. 4.1 Electrode placement around the forearm. First channel pair is placed along the extensor carpi ulnaris with the rest equally spread around the forearm with the positive lead as the most proximal one.

but with MVC force. When obtaining minimum and maximum contraction forces each movement was performed only one time and for 3 seconds with 6 seconds resting between each movement. The MVC force was calculated by removing 20% from the start and end of the signal from each channel and computing the mean absolute value (MAV) of each channel in 100 ms time windows. Then the mean of all the channels MAVs was calculated. This was done for every movement. The MVC force was used to display a ramp from 0 to 100% (lowest to highest contraction force) for each movement. A red marker was displayed in the same plot as the ramp showing the current contraction force, see figure 4.3. The current contraction force was calculated by taking the MAV of the samples from a time window of 100 ms as the recording was performed. The movements were then recorded in the same way as for the standard recording session with the only difference that the subjects were asked to track the ramp with the red marker during the recording of each movement. For the repetition of the movements the red marker was reset to 0 in the lower left corner.

Each subject had to perform at least four test ramp recording session to let the subjects familiarize with the ramp recording session. Some of the subjects did the test recording session up to seven and eight times to provide acceptable ramp recordings. In order to determine if the recorded ramp data was acceptable, the data was visually inspected after each ramp recording session. The recorded signal from one channel that was seen as an example of an acceptable recording is shown in figure 4.4. The signal in figure 4.4 was used as a reference for what the other subjects should aim for. The recorded MES in the figure starts with low amplitude and is increasing throughout the recording as a ramp. It should be noted that the recorded signals for certain movements could have acceptable ramp signals for some channels but not for another channel, mainly due to that the muscles under some electrodes are more active when performing some movements. Also some movements like flex hand and supination were hard to do in a way that generated acceptable ramp recordings for most subjects.

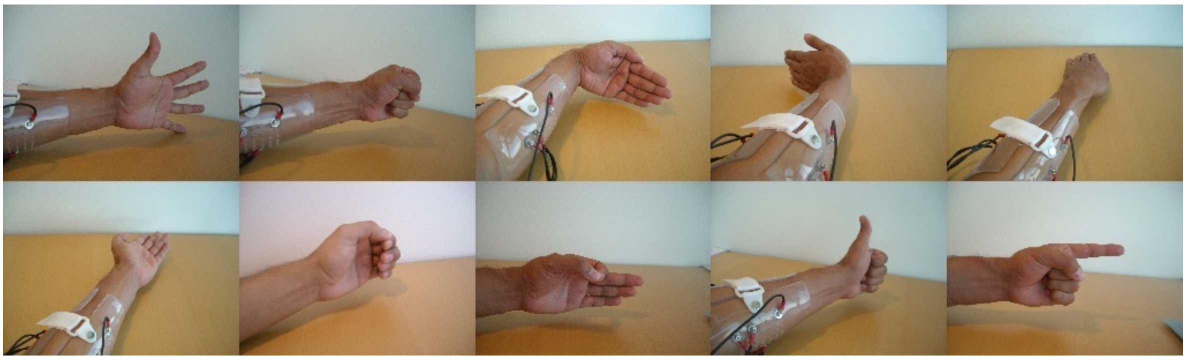


Fig. 4.2 Different movements. Starting from upper left and going to the right: Open hand, Close hand, Flex hand, Extend hand, Pronation, Supination, Side grip, Fine grip, Agree and Pointer.

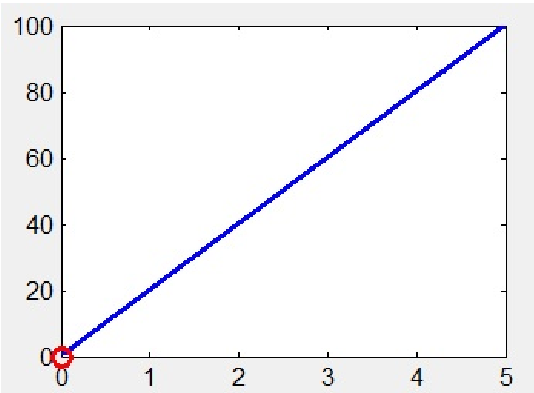


Fig. 4.3 Ramp and tracer with the ramp ranging from 0 to 100% of maximum voluntary contraction force over 5 seconds.

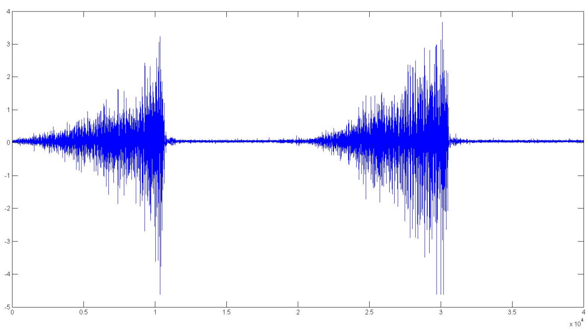


Fig. 4.4 Signal from one of the channels in a recording that was viewed as high quality.

4.5 Offline accuracy

The offline accuracy has been evaluated in order to compare different classifiers performance, trained with ramp and sustained data, and with different settings (topology and feature set) affecting how the classifiers are trained. It should be noted that the data used when comparing ramp and non-ramp accuracies were recorded at the same day and in close sequence so that each subject had the same electrode placement when doing both recording sessions. This is advantageous when comparing the ramp and non-ramp accuracies to be sure that the electrode placement is not affecting the results. The offline evaluation has been performed two times. First one time as a pre-study to find suitable combination of settings to use for the real-time evaluations and a second time (main study) using the data that was used to train the classifiers for the real-time accuracies (the motion tests described later). By using the same data for training of classifiers to evaluate real-time accuracies and for offline accuracies the relation between real-time and offline can be investigated, even if the relationship has shown to be vague. For the pre-study five subjects participated, and for the main study six subjects. The pre-study offline tests were performed using ramp data that was not of as high quality as the data used for the main study, mainly due to inexperience with the ramp recording session. However the results obtained from the pre-study were only used as a guidance when performing the main study including the motions tests. As mentioned, the same data was used for the main offline accuracies and for training the classifiers for motions tests. The offline accuracies were calculated when all the motions test were done. The results for the pre-study are not presented since no conclusion are made from those results. The main offline study and its results are presented in the results section. As previously mentioned, the offline tests were performed to find combinations of classifier, topology and feature set that would be used for motion tests. The classifiers that were compared are linear discriminant analysis (LDA), backpropagated multilayer perceptron (MLP) and mean Regulatory Feedback Networks (RFN). RFN was only used for the offline accuracy, and not for motion tests, since this algorithm resulted in considerably lower average accuracy compared to LDA and MLP. The different settings that have been evaluated in the offline tests are different topologies and feature sets. The topologies that were tested are Single, One vs. One and One vs. All (for definition of the different topologies see [19]). The feature sets that were tested are f1 = tmabs, twl, tzc, tslpch, f2 = tmabs, fwl, trms, tzc and f3 = tstd, fwl, fmd (for full name of the features see the list of abbreviations). These features were used because they were shown to be promising in a study comparing feature sets [20]. In order to find a combination of classifier, topology and feature set that would be used for motion tests, the different setting have been changed one at a time. First the accuracy between ramp and non-ramp trained classifiers was evaluated. Next the performance for ramp trained classifiers were evaluated with different topologies and one feature set (f1). Last the ramp trained classifiers were evaluated when trained with different feature sets and topology set to one vs. one since this topology was found to be the best in the previous step. When the data was recorded for the pre-study with ramp and sustained contractions, three of the five subjects started with the standard recording session and two subjects started with the ramp recording session in order to minimize any unwanted effects on the results due to that all of the subjects started with the same session. For the main study five of six subjects started with the ramp recording session. The reason was that most subjects had to perform the ramp recording session several times, at least four times, before recording the real ramp recording session. Some subjects had problems generating acceptable ramp recordings so when the subjects did generate acceptable data that data was used. The offline studies were performed in the following way for both the pre-study and the main offline study:

1. As the first step, the accuracy between classifiers trained with ramp and non-ramp data was compared. For this comparison, the classifiers used were LDA, MLP and RFN. The topology was set to one vs. one. The feature set used was f1 = {tmabs, twl, tzc, psltch2}. For each subject and for both the ramp and sustained contraction data the offline accuracy was calculated 100 times in the pre-study, but only 10 times in the main study. The accuracy was calculated only 10 times in the main study to reduce the time it took to get all the accuracies. For example it could take more than 24 hours to calculate the accuracy 100 times when training the classifier with MLP. Comparing the cases when 10 or 100 repetitions were used for MLP using single topology and feature set f1, the average accuracy was 91.8% and 91.5% for 10 and 100 repetitions respectively. The standard deviation was 5.2 for both 10 and 100 repetitions respectively.

2. The main test was done to compare different topologies. For this test LDA, MLP and RFN were used together with feature set f1 as in the first step. The following topologies were compared: Single, One

vs. One and One vs. All. For this test the offline accuracy for each case and subject was calculated 10 times for both the pre-study and the main offline tests.

3. The third test was done to compare different feature sets. For this test LDA, MLP and RFN were trained with topology set to one vs. one, since this topology was found to yield the best average accuracies in the previous test. The feature sets used were f1 = {tmabs, twl, tzc, tslpch}, f2 = {tmabs, fwl, trms, tzc} and f3 = {tstd, fwl, fmd}. As in the previous step the accuracies were calculated 10 times for both the pre-study and the main offline tests.

The results from all the tests were analyzed using the average accuracy for all ten movements and average accuracy of two subsets of the ten movements. The classifiers were not trained specifically with the movements in the subset of movements shown below. The eleventh movements which is rest was excluded from all the average accuracies. The reason for dividing the movements into subsets was to evaluate if there is any difference in average accuracy depending on the set of movements.

- **10 movements:** {Open hand, Close hand, Flex hand, Extend hand, Pronation, Supination, Side grip, Fine grip, Agree and Pointer}
- **6 movements:** {Open hand, Close hand, Flex hand, Extend hand, Pronation, Supination}
- **4 movements:** {Side grip, Fine grip, Agree and Pointer}

The accuracy for the sets of movements and a case in the tests were analyzed by plotting error bars and calculating the average accuracy for the cases compared in each test. How the average accuracies were obtained is explain below step by step:

1. The average over the repetitions of each movement from the data of every subject was computed. This resulted in a vector with 10 average values, one for each movement, for every subject.
2. Depending on the set of movements presented above (10, 6 or 4 movements), these were picked out from the total number of movements.
3. The six subjects' vectors with average values were then put into one new vector. This was done for every set of movements (10, 6 or 4 movements). This resulted in one vector with length of 60, 36 or 24 values depending on the set of movements.
4. The vectors containing all the subject' average accuracies were then used in a boxplot for each case compared. The data was also used to obtain a mean value which represented the total average accuracy.
5. The total average accuracies were then compared against the other cases average accuracy. For example in the first test the classifiers trained with ramp data was compared with the classifiers trained with sustained contraction data.

4.6 Real time accuracy

Even though the relation between offline accuracy and the real-time accuracy is vague [21], the number of classifiers, features and topologies to be compared had to be limited, and therefore the result from the pre-study were used to know which combinations to compare in the real-time tests. For the real-time tests each subject performed three motion tests, one for each of the setting in table 4.1 (the motion test is described later).

Table 4.1: Setting used for performing motion tests

Setup	Data	Classifier	Topology	Feature set
1	Sustained contraction	LDA	One vs. One	f1{tmabs, twl, tzc, tslpch}
2	Ramp contraction	LDA	One vs. One	f1{tmabs, twl, tzc, tslpch}
2	Ramp contraction	MLP	One vs. One	f1{tmabs, twl, tzc, tslpch}

The subjects started with either a sustained or a ramp recording session followed by a motion test. As mentioned five out of six subjects started with the ramp recording session. The two motion tests performed with ramp data, where the classifiers were trained with LDA or MLP, were always done in sequence. It was desirable to use the same data to train both the LDA and the MLP classifier for the ramp case so that the time after the data was recorded is about the same. Also, using the same data ensures that there is no difference in quality of the data used to train the classifiers.

Three subjects did the motion tests in the order: setup 2, setup 3 and setup 1. Two subjects did them in the following order: setup 3, setup 2 and setup 1. One subject did it in the order: setup 1, setup 3 and setup 2.

4.7 Preprocessing of data and signal processing

The data recorded from the ramp and sustained recording sessions had to be preprocessed in order to use it for classifier training. The preprocessing of the data started by removing 15% of the samples from the beginning and the end of the recorded data for each channel, movement and for each repetition of each movement. The 15% of the beginning and end of the signals was removed because the subjects do not normally start or stop the contractions precisely when instructed to. The remaining data was divided into time windows of 200 ms with an overlap of 50 ms between each window. These windows were then divided into training data (50% of windows), validation data (10%) and testing data (40%). This partitioning used the first 50% of the data recorded to train the classifier, which corresponds to one ramp (the first ramp out of the two repetitions). Partitioning the data in this way assures that the data used to train the classifiers is represented by one whole ramp and does not contain more information from any specific effort level.

When the training of the classifier is performed, the time windows chosen for training, validation and test are randomized. The sets used for training were romanized because this resulted in much better accuracy. In figure 4.5, two confusion matrices are shown, with dandomized sets and one without randomization. The axis in the figures represent the movements. The colors in the squares represent the accuracy for that combination, where red is high accuracy and blue is low. The confusion matrix is obtained by cross validation between all movements. Ideally every element should me blue except those on the diagonal which should be red.

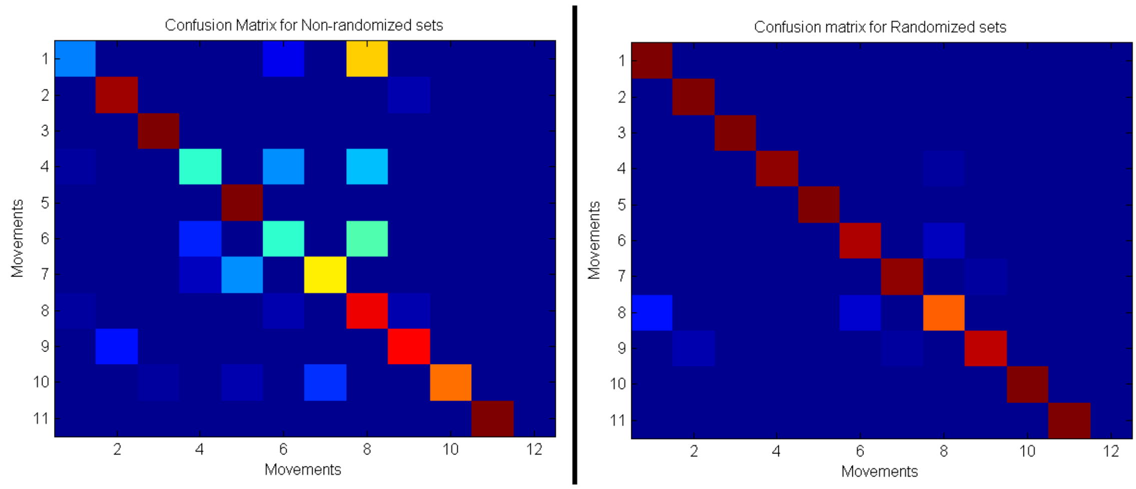


Fig. 4.5 Difference in classification accuracy when comparing randomized and non-randomized sets for training of classifiers. In this example an LDA classifier is used with one vs. one and feature set f1. A perfect confusion matrix would have all blue squares except for the diagonal which should have all red squares.

4.8 Motion test

Motion tests are used to evaluate the real-time accuracy of the movements. For each trial the subjects is instructed to perform the movements that are displayed. The movements are displayed in a random order where each movements is displayed three times. This was done in two consecutive sessions. If the subjects does not manage to get 20 correct predictions within ten seconds, that movements is incomplete and the test moves on to the next movement. Between each movement a resting period, of about three seconds depending on the computer hardware, is included. This protocol is repeated as many times as specified by the number of trials. The motion tests were carried out in two trails, three repetitions and ten seconds time out. In total the movements were performed six times. More details on the motions tests can be found in [18].

4.9 Statistical significance

In order to validate if the results obtained from comparing the different accuracies are statically significant, the Wilcoxon signed rank test was used. This test was chosen because Demšar J. compared different tests and found that when comparing one classifiers against another, that is trained with the same data, the Wilcoxon signed rank test was the most suitable one to use [22]. In this thesis the Wilcoxon signed rank test is used to compare classifiers trained with different topologies and feature sets, using the same data. The results were assumed to be statistically significant at $p<0.05$.

4.10 Boxplots

Boxplots were used to show the data obtained. The horizontal bar in the boxes represents the median value of the data set and the arrows represent the mean value. The edges of the box reaches from the 25^{th} to the 75^{th} percentile. Lower whiskers extend to $q_3 + 1.5(q_3 - q_1)$ and upper whiskers extend to $q_1 - 1.5(q_3 - q_1)$, where q_1 and q_3 are the 25^{th} and 75^{th} percentiles respectively. Data points outside the whiskers are outliers and are represented by ”o” signs.

Chapter 5

Results

5.1 Offline accuracy

The results from comparing the LDA, MLP and RFN classifiers trained with ramp and sustained contraction data are shown in figure 5.1, 5.2 and 5.3 for the different cases ten, six and four movements respectively. All the accuracies were calculated using feature set f1 and topology set to one vs. one. Results show that for every classifier and set of movements (10, 6 and 4 movements), the average accuracy is better when the classifier is trained with the non-ramp data ($p<0.05$).

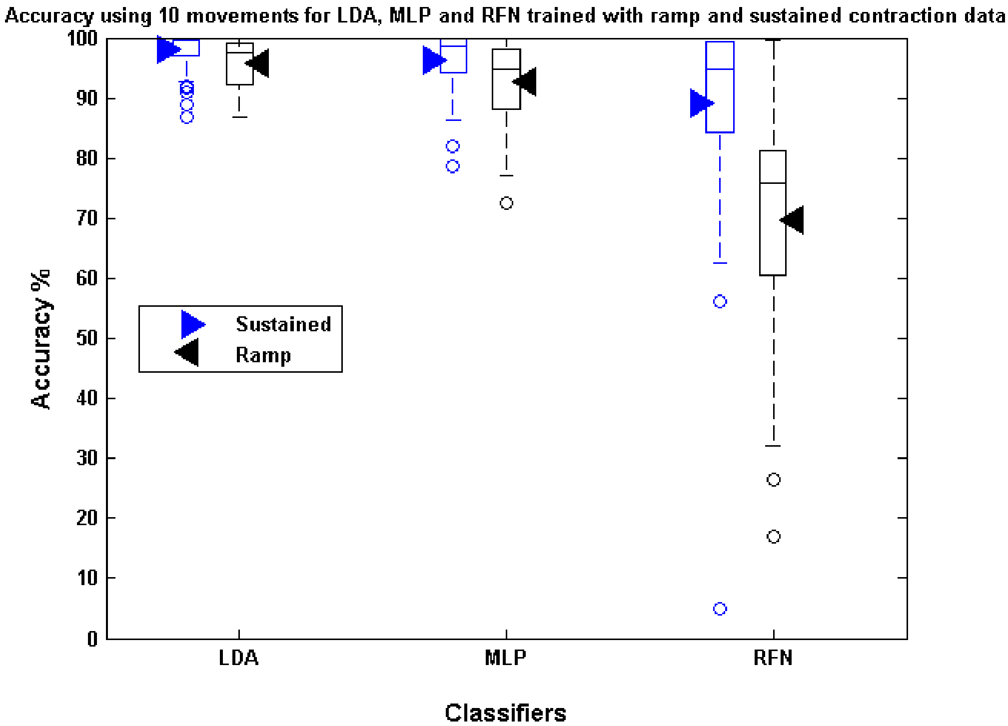


Fig. 5.1 Average of 10 movements' offline accuracy for LDA, MLP and RFN when the classifier was trained with ramp and sustained contraction data. The feature set used was set f1 and the topology used was one vs. one. Arrows represent mean values.

The plots from comparing the average classification accuracy when the classifiers were trained with different topologies are shown in figure 5.4, 5.5 and 5.6 for ten, six and four movements respectively. For this experiment feature set f1 was used. The results suggests that one vs. one results in better offline accuracy than single and one vs. all for most classifier and for all the set of movements. Two exception are found in the case of ten and six movements using RFN, however the accuracy is also lower for RFN than for LDA and MLP ($p<0.05$). With the Wilcoxon signed rank test the results for comparing the accuracies using different

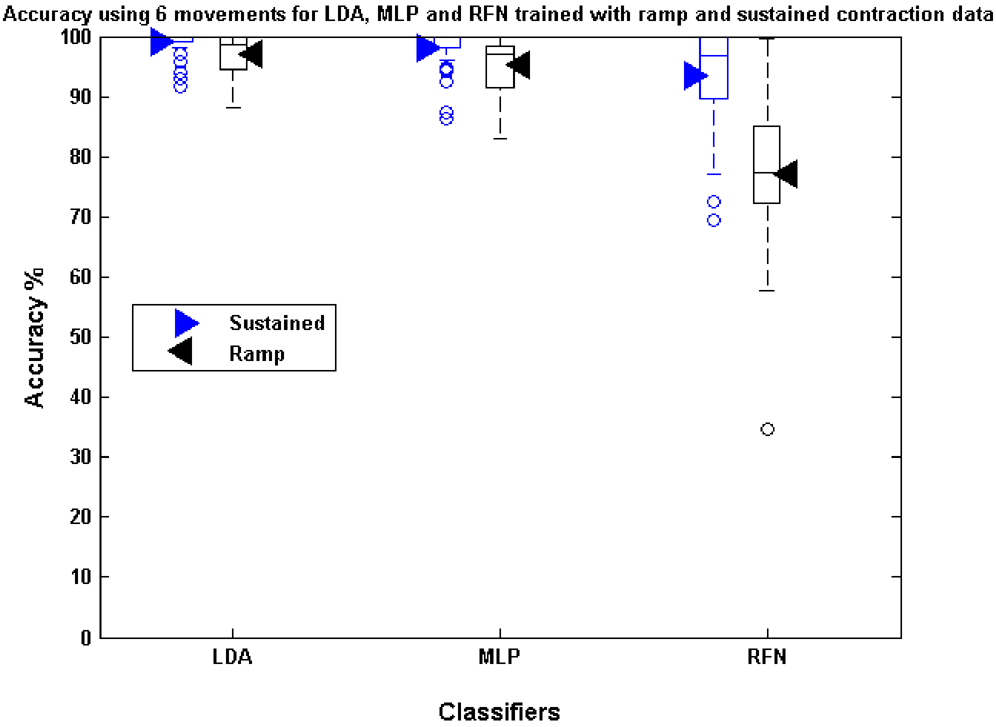


Fig. 5.2 Average of 6 movements' offline accuracy for LDA, MLP and RFN when the classifier was trained with ramp and sustained contraction data. The feature set used was set f1 and the topology used was one vs. one. Arrows represent mean values.

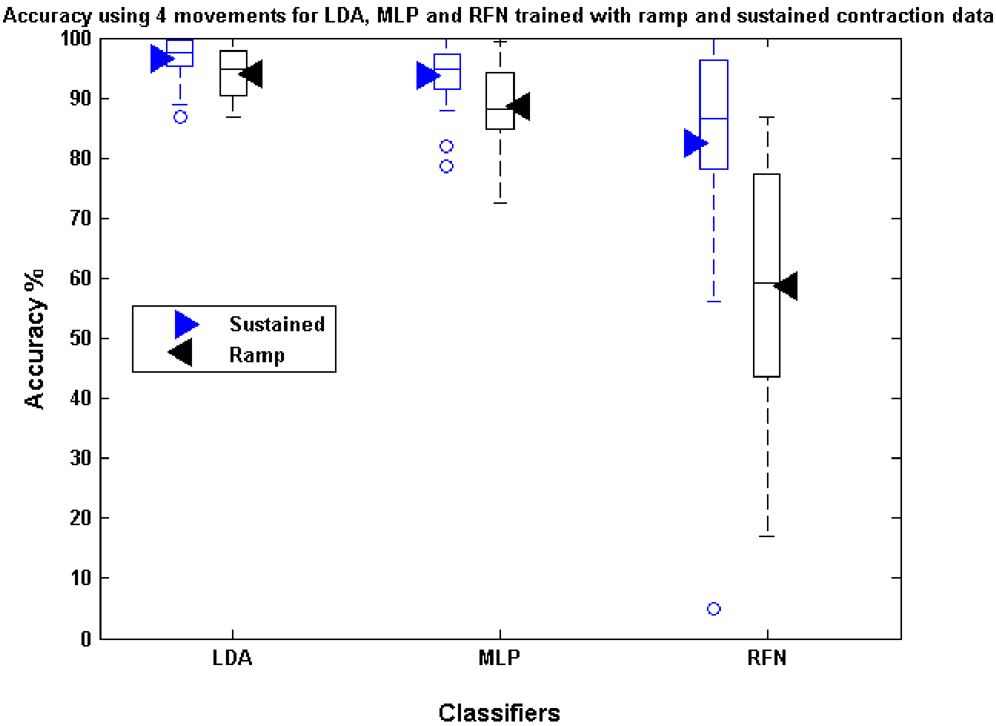


Fig. 5.3 Average of 4 movements' offline accuracy for LDA, MLP and RFN when the classifier was trained with ramp and sustained contraction data. The feature set used was set f1 and the topology used was one vs. one. Arrows represent mean values.

topologies for the same classifier were found to be statistically significant (p<0.05) for all combination except two. The first exceptions was for ten movements, RFN, comparing one vs. one and single (p=0.59). The other exception was for four movements, MLP comparing single and one vs all (p=0.14).

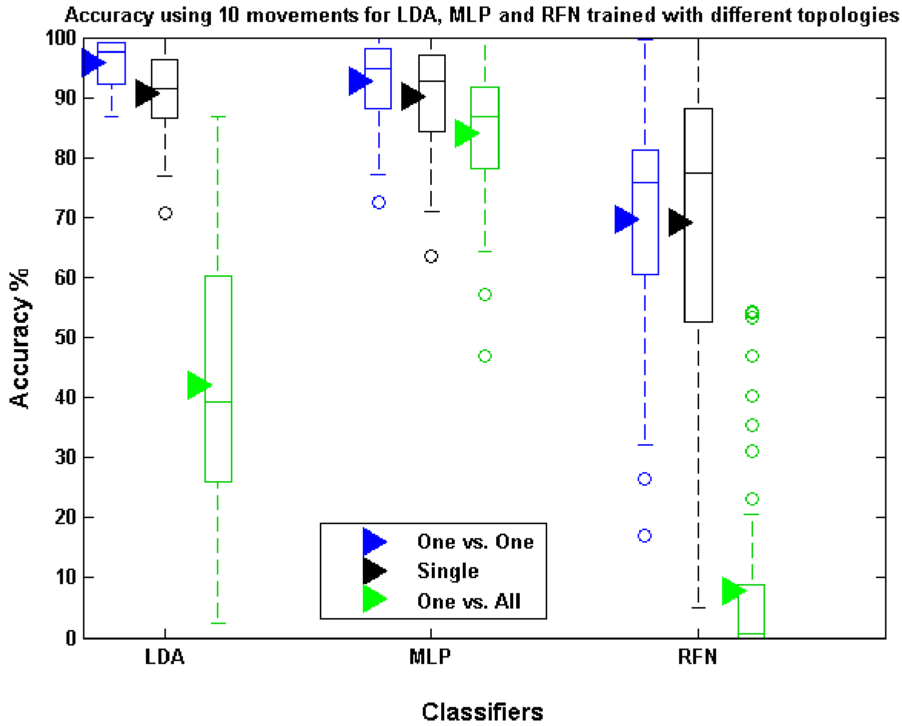


Fig. 5.4 Average of 10 movements' offline accuracy for LDA, MLP and RFN when trained with ramp data and different topologies. Feature set f1 was used. Arrows represent mean values.

The plots from comparing the average classification accuracy for the classifiers when trained with topology set to one vs. one and different feature sets (f1, f2 and f3) are shown in figure 5.7, 5.8 and 5.9. All comparisons are statistically significant (p<0.05). For RFN the best feature set varies depending on the set of movements, however the accuracy for RFN is also lower for every feature set compared to LDA and MLP for the same feature set (p<0.05). With this arguments the best feature set is f1.

From the offline study the best setups are LDA and MLP trained with feature set f1 = {tmabs, twl, tzc, tslpch} and topology set to one vs. one.

5.2 Real-time accuracy

The accuracies form the motion tests are shown in figure 5.10, 5.11 and 5.12. From the results in the figures it is not possible to say that of the three cases results in the best average accuracy. Looking at the figures the case that yields best average accuracy depends on which set of movements that is used. When looking at all ten movements, setup 1 gives an average accuracy of 63.32% and setup 2 gives 62.46% and setup 3 gives 63.03%. These results for setup 1 and setup 2 are very close and the significance test gives p = 0.549, which means that the difference between the two cases is not statistically significant. Setup 1 and setup 2 both gives higher average accuracies than setup 3 (p<0.05).

Looking at the six movements, setup 1 gives an average accuracy of 68.59%, setup 2 gives 63.89% and setup 3 gives 63.03%. However, the difference between setup 1 and setup 2 is not significant (p = 0.22).

Looking at the four movement case, setup 1 gives an average accuracy of 55.43%, setup 2 gives 60.00% and setup 3 gives 45.25%. Setup 2 gives the best average accuracy, but the difference between setup 1 and setup 2 is not statistically significant (p = 0.72) and the difference between setup 1 and setup 3 is not significant (p = 0.13).

Comparing the classifiers trained with ramp data the better classifier is LDA. This is statistically significant except for six movements.

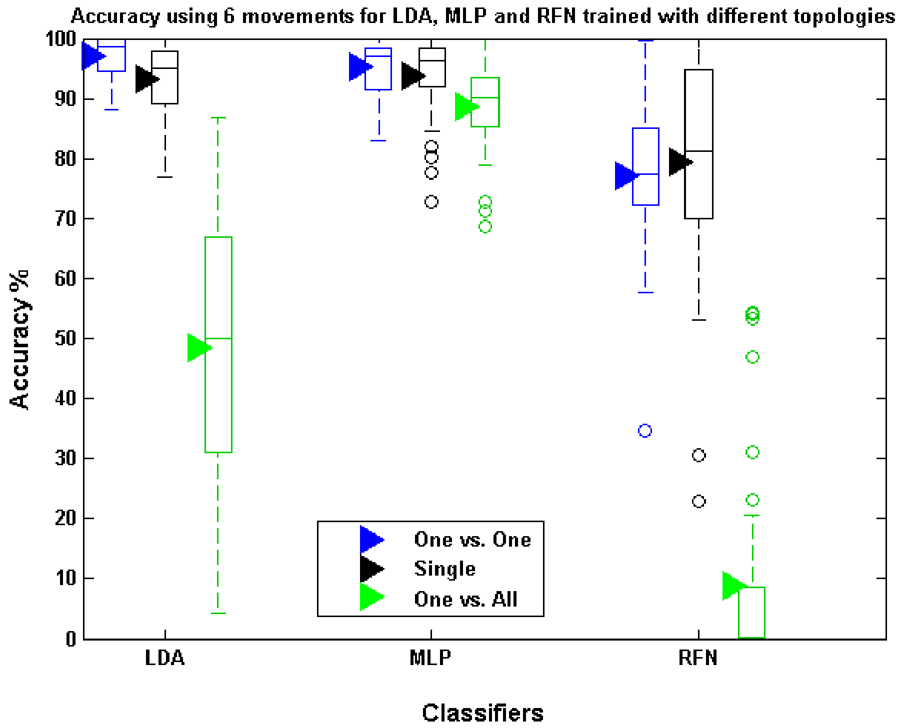


Fig. 5.5 Average of 6 movements' offline accuracy for LDA, MLP and RFN when trained with ramp data and different topologies. Feature set f1 was used. Arrows represent mean values.

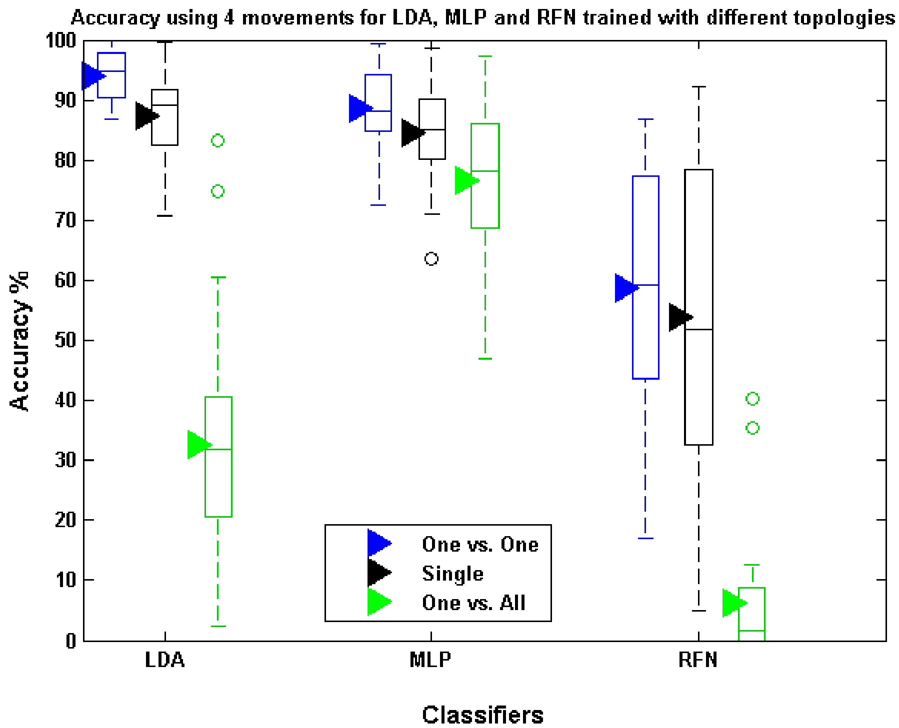


Fig. 5.6 Average of 4 movements' offline accuracy for LDA, MLP and RFN when trained with ramp data and different topologies. Feature set f1 was used. Arrows represent mean values.

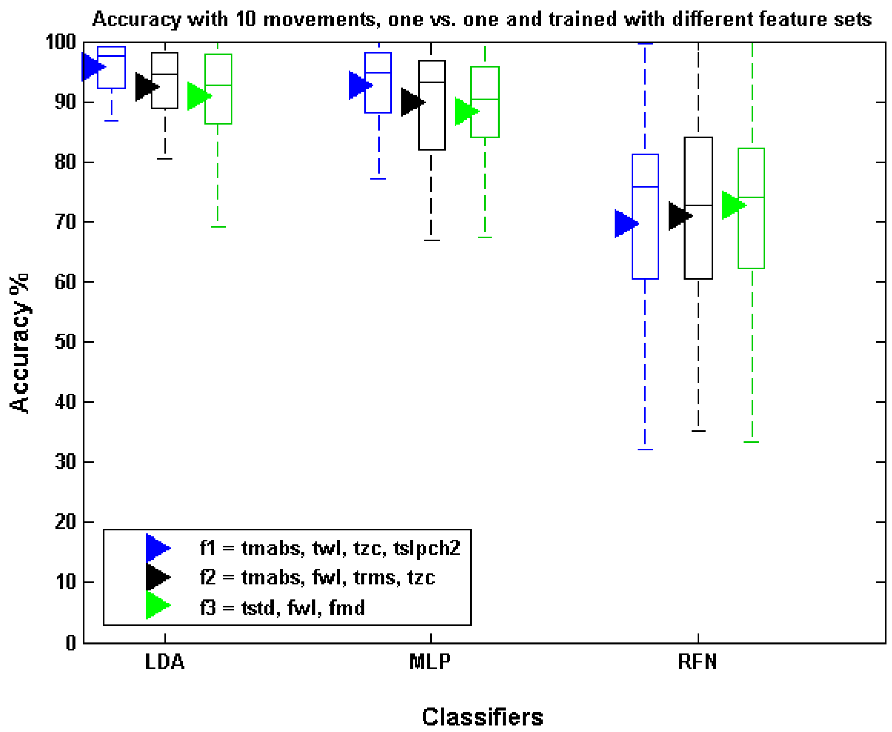


Fig. 5.7 Average of 10 movements' offline accuracy for LDA, MLP and RFN when trained with ramp data and different feature sets. Topology was set to one vs. one. Arrows represent mean values.

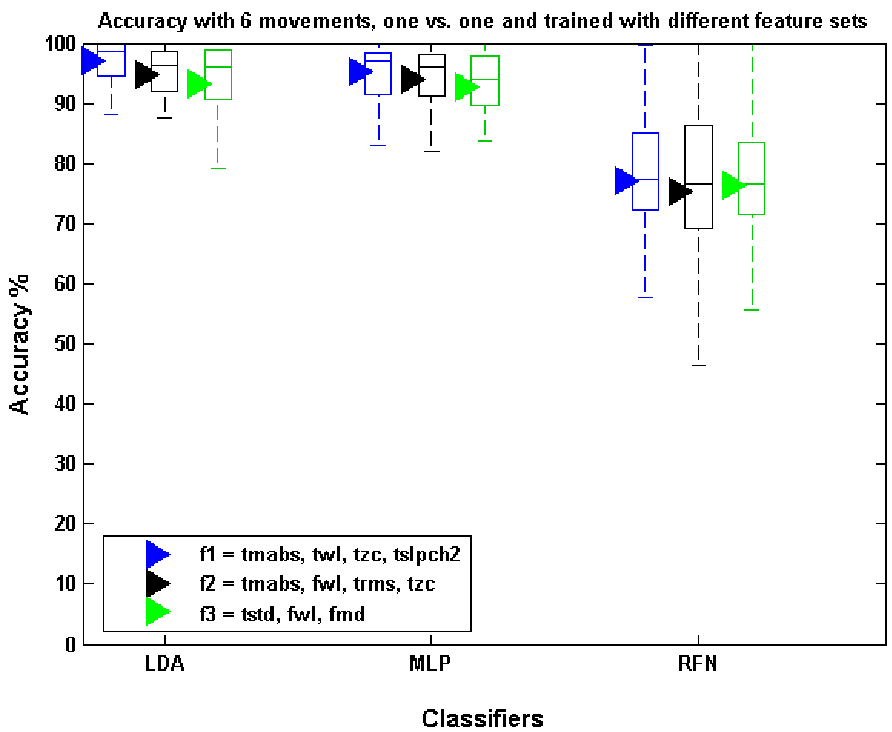


Fig. 5.8 Average of 6 movements' offline accuracy for LDA, MLP and RFN when trained with ramp data and different feature sets. Topology was set to one vs. one. Arrows represent mean values.

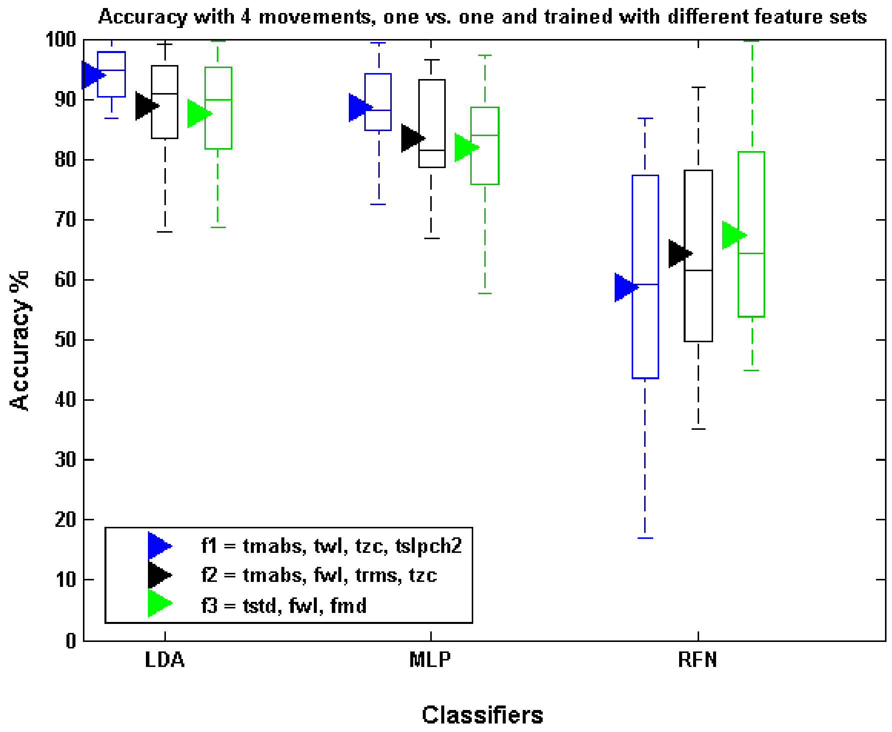


Fig. 5.9 Average of 4 movements' offline accuracy for LDA, MLP and RFN when trained with ramp data and different feature sets. Topology was set to one vs. one. Arrows represent mean values.

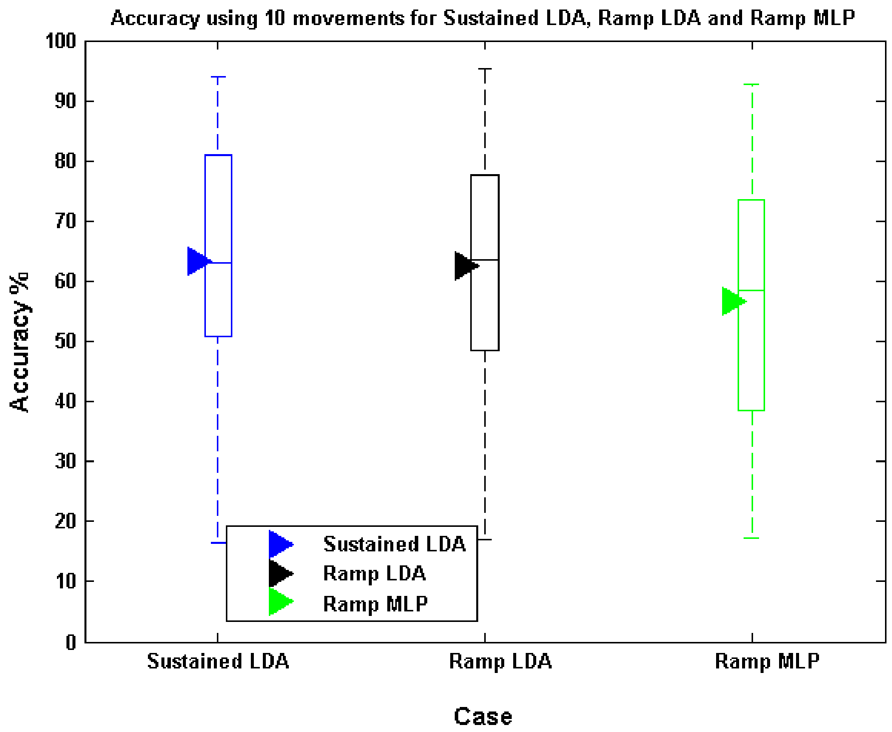


Fig. 5.10 Average of 10 movements' real-time accuracy using motion tests. For all three cases feature set f1 has been used with topology set to one vs. one. The cases that have been compared are when training the classifier with sustained data using LDA, ramp data using LDA and ramp data using MLP. Arrows represent mean values.

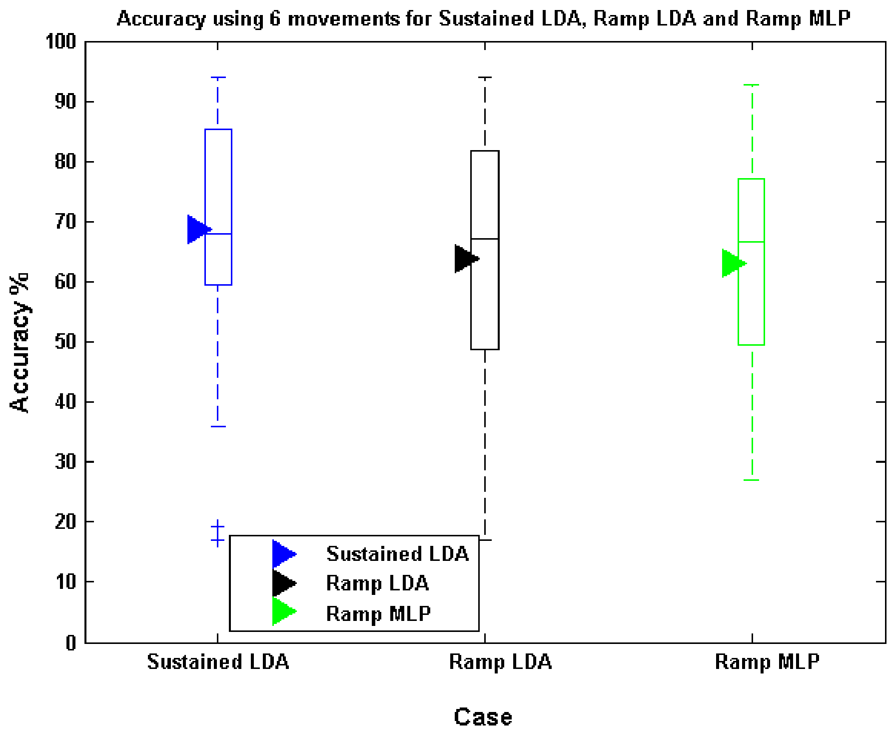


Fig. 5.11 Average of 6 movements' real-time accuracy using motion tests. For all three cases feature set f1 has been used with topology set to one vs. one. The cases that have been compared are when training the classifier with sustained data using LDA, ramp data using LDA and ramp data using MLP. Arrows represent mean values.

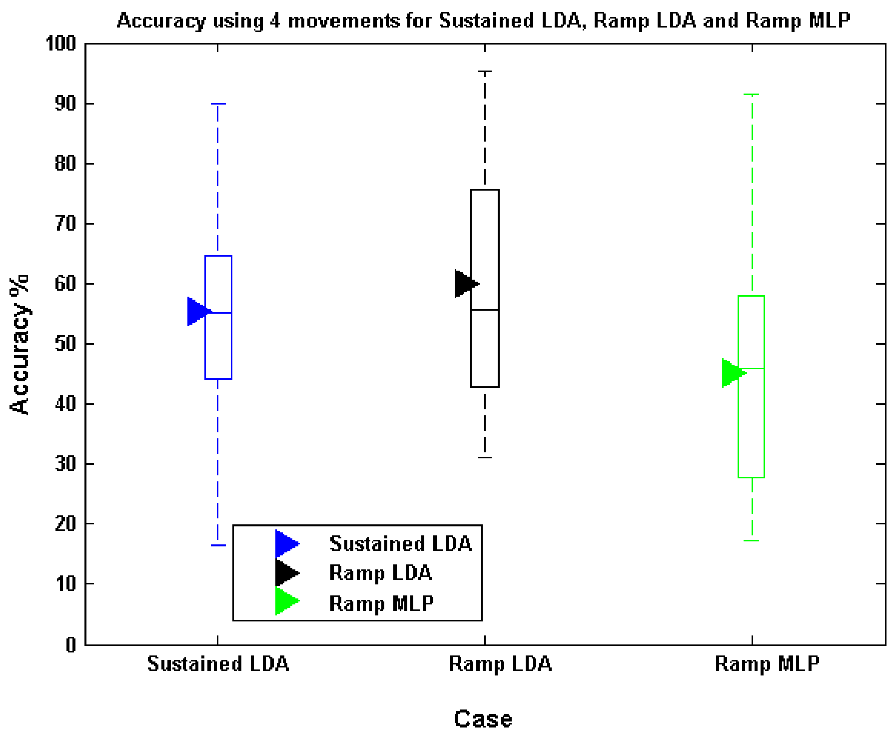


Fig. 5.12 Average of 4 movements' real-time accuracy using motion tests. For all three cases feature set f1 has been used with topology set to one vs. one. The cases that have been compared are when training the classifier with sustained data using LDA, ramp data using LDA and ramp data using MLP. Arrows represent mean values.

5.3 Real-time and offline accuracy

The comparisons between the real-time accuracy and the offline accuracy are shown in figure 5.13, 5.14 and 5.15. The graphs shows that the offline accuracy is much higher than the real-time accuracy for every case ($p<0.05$). For the offline case the bast accuracy is obtained from setup 1, second best is setup 2 and lowest accuracy is obtained from setup 3 ($p<0.05$). For the real-time accuracies the best setup is not always setup 1, then setup 2 and last setup 3 as for the offline accuracies. Instead the best setup depends on the set of movements.

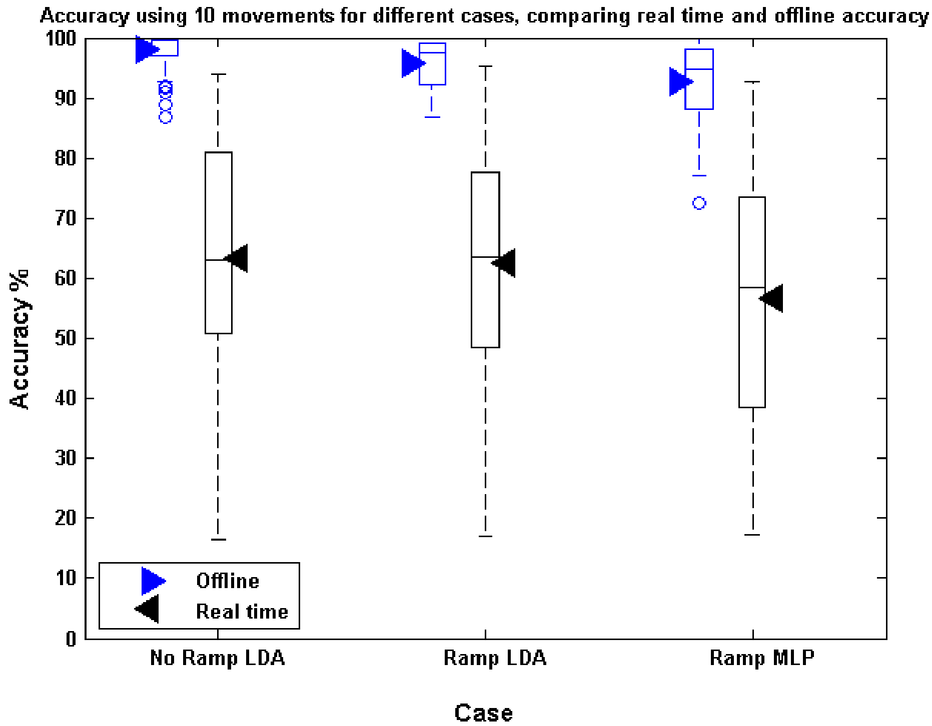


Fig. 5.13 Average of 10 movements' real-time accuracy and offline accuracy. For all three cases feature set f1 has been used with topology set to one vs. one. The cases that have been compared are when training the classifier with sustained data using LDA, training with ramp data using LDA and training with ramp data using MLP. Arrows represent mean values.

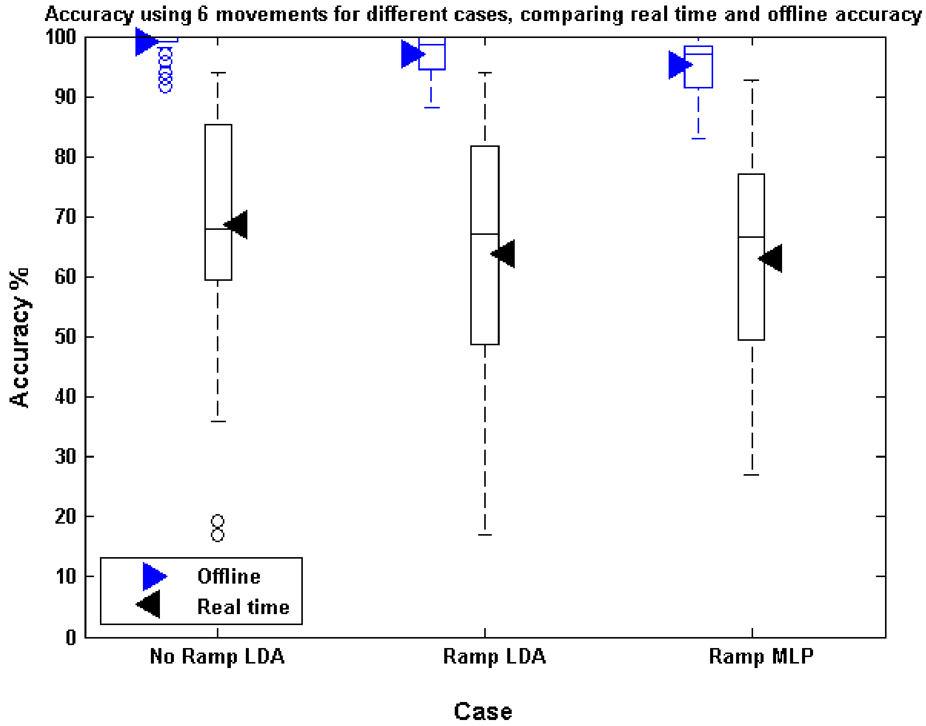


Fig. 5.14 Average of 6 movements' real-time accuracy and offline accuracy. For all three cases feature set f1 has been used with topology set to one vs. one. The cases that have been compared are when training the classifier with sustained data using LDA, training with ramp data using LDA and training with ramp data using MLP. Arrows represent mean values.

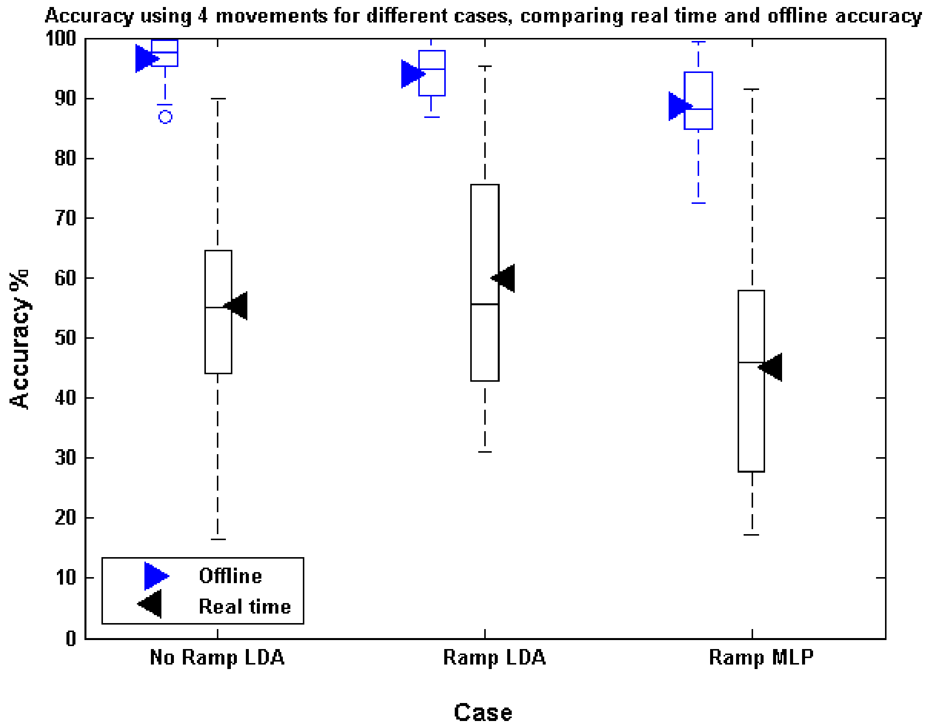


Fig. 5.15 Average of 4 movements' real-time accuracy and offline accuracy. For all three cases feature set f1 has been used with topology set to one vs. one. The cases that have been compared are when training the classifier with sustained data using LDA, training with ramp data using LDA and training with ramp data using MLP. Arrows represent mean values.

Chapter 6

Discussion

It should be noted that even though different set of movements were compared (10, 6 and 4 movements) the accuracies obtain from the tests are from using all ten movements when training the classifiers. The different sets of movement are only different in the movements used to compute the average accuracy of the movements. If the classifier would actually be trained with ten, six and four movements the average accuracies might have been higher when only using six and four movements since the classifier would not have to distinguish between as many movements and thereby be more accurate.

6.1 Recording sessions

When the MES for the movement were recorded each repetition of each movement was a contraction of five seconds. Five seconds was used since it would be hard to follow the ramp if the contraction time is shorter. The drawback with performing five second contractions could be that the subjects experience fatigue when using this long contraction time, especially when doing the ramp recordings since these involve hard contractions. However it was important to make it easier for the subjects to produce acceptable ramp signals, and therefore five seconds was used. Longer than five seconds could have been too exhausting for the subjects and lowered their ability to produce acceptable ramp data. Five second contractions was also used for the sustained recording session so that the conditions would be the same for ramp and sustained recordings.

When performing the ramp recording session the subjects' ability to produce data increasing as a ramp varied. It was noticed that one of the subjects that had done the ramp recording sessions many more times than the other subjects was able to produce recording that was more ramp looking. Because of this it would be ideal to let all the subjects perform the ramp recording session as many times as possible in order for them to be able to produce very high quality ramp data. However, time was a problem since the subjects could not spare a lot of time to learn to produce perfect ramp data. The reason for making sure that the subjects could produce acceptable ramp data was that there had to be a difference between the data compared, that is that there is a difference in the ramp and sustained contraction data. If the difference is not very big, which could be the case if the ramp recording are not acceptable, there is no point in comparing the data since the comparison would actually not show upon a difference between training classifiers with ramp or sustained contraction data.

6.2 Offline accuracies

The method used when finding combinations of classifier, topology and feature set is possibly not the most solid approach. To be sure that the optimal combination is found an alternative would be to train every classifier with every topology and with every feature set and then compare every combination. However, this could not be done since this would result in to many combination to test. Because of this, the method used where that only one setting was changed and where the combination with the best average accuracy was assumed to be the best, even though this setting had not been tested again every feature set and topology. A possible way of doing this better would be to automate the process of running the offline accuracies so that all possible combinations could be tested one after another. However the MATLAB program used to calculate the offline accuracies did not allow for this to be done and would have to be changed in order to do so, which would be very time consuming.

The results obtained from comparing the offline accuracy were that the accuracy was better when training with sustained data. When the offline accuracy is calculated parts of the recorded data set is used to train the classifiers, and parts are used to test and calculate the accuracy. It is probable that data used for training and validating the classifier is more similar for the sustained contraction data than for the ramp data since it is easier to get similar recordings from the two repetitions when performing a steady effort contractions. The reason for obtaining high offline accuracies with the sustained contraction data is that the data used to train and test the classifier is from the same contraction level and thereby more similar than the ramp data.

6.3 Real-time accuracy

The results obtained from the motion tests suggests that training the classifier with sustained contraction data and ramp data does not make a statistically significant difference. When performing the motion tests the subjects were asked to perform the movements without any specific instructions on how hard the contraction should be. The subjects performed the movements with an intermediate contraction force, similar to the force used for the sustained contraction recordings. Despite this there is no significant difference. The result that training with ramp and testing with sustained ”normal” contraction force does not lower the average accuracy statistical significantly implies that it could be that it is not worse to use ramp data. Results by Scheme et al. [15] showed that ”training with the dynamic ramp data significantly improved ($p<0.001$) both classification error ($14.84 \pm 0.60\%$ and $11.16 \pm 0.54\%$ for medium and ramp training, respectively) and R2 tracking score (0.146 ± 0.037 and 0.328 ± 0.028 for medium and ramp training, respectively) during the tracking task”. With this result it should be taken into consideration that to track a sinusoidal curve with varying frequency different effort levels has to be used in order to follow the curve. The motion test performed in this thesis tested the real-time accuracy when trying to achieve the correct movement with sustained ”normal” contraction force. Since the results from this thesis showed that the average accuracy when training with ramp and testing with sustained ”normal” force did not decrease, and the results from Scheme et al. improved performance when testing with many different effort levels, it is likely that training with ramp data is overall better for accuracy.

6.4 Offline and real-time

The offline accuracy results from Scheme *et al.* [15] and Al-Timemy *et al.* [16] shows the offline error rate increases when training with ramp data or data from several effort levels and testing with sustained ”normal” contraction, compared to training and testing with sustained ”normal”. These results are compliant with the offline result from the study in this thesis. However, the results obtained from the study in this thesis suggests that the real-time accuracy is similar when training with ramp and non-ramp data. The reason that the average accuracy is relatively high in the real-time study might be that when performing a motion test it can be hard to perform the movements with the same effort as when recording the training data. However, when training with ramp data there is more information from many effort levels and thereby it may be easier, or at least not harder to perform the correct movement. This also suggests that there is no obvious connection between offline and real-time accuracy as found is other papers.

6.5 Statistical significance

The Wilcoxon signed rank test was used when evaluation the statistical significance for the accuracies calculated in the offline evaluation and in the real-time evaluation. For some comparisons the difference was found not to be statistically significant and thereby no conclusion could be made about which of the cases or setups compared that was the best. If the data is similar and if the difference is not significantly different it is possible, but not a fact, that the things that are compared could generate similar average accuracy.

6.6 Future work

If further studies into this area would be performed a key aspect to focus on could be to make sure that the subjects are experienced with the ramp recording session and can thereby produce high quality recordings. It could be of interest to investigate the relation between experience with the ramp recording and the real-time accuracy in order to see if there is anything to gain by more experience. Since some of the setups or cases in this study were not statistically significant it would be of interest to perform similar experiments with more subjects to increase the chance of obtaining significant comparisons. It could also be of interest to investigate real-time accuracy when training with ramp and performing motion tests with different effort levels and compare results. Since no classifier was trained for classifying only the subset of movements (4 and 6 movements) it would be of interest to do this.

Chapter 7

Conclusions

Results from the main offline study were that LDA and MLP with topology set to one vs. one and feature set f1 was best out of the combinations tested.

The results from the real-time accuracy, using the setups from the *Real-time accuracy* subsection in the method, does not show a significant difference between setup 1 and setup 2 and therefore it is not possible to determine if one is better than the other. Setup 1 and setup 2 do however give higher average accuracy than setup 3 for ten and six movements, but for four movements the difference between setup 1 and setup 3 is not statistically significant. Because of this it is probable that setup 1 and setup 2 give similar accuracy for real-time pattern recognition. It is also possible that when looking at four movements, setup 1 and setup 3 give similar average accuracy. Setup 2 gives better average accuracy than setup 3 for all set of movements which means that if ramp data is used, LDA is better to use than MLP.

In the offline study the results show that the average offline accuracy is better when training with sustained contraction data. However, since the relation between using ramp and sustained data is not the same in real-time, there is no direct or obvious connection between offline and real-time accuracy.

References

[1] N. Jiang, S. Dosen, and D. Farina, “Myoelectric Control of Artificial Limbs - Is There a Need to Change Focus?” *in the SPOTLIGHT*, no. SEPTEMBER, pp. 12–15, 2012.

[2] A. Fougner, O. y. Stavdahl, P. J. Kyberd, Y. G. Losier, P. A. Parker, and S. Member, “Control of Upper Limb Prostheses : Terminology and Proportional Myoelectric Control - A Review,” *IEEE TRANSACTIONS ON NEURAL SYSTEMS AND REHABILITATION ENGINEERING*, vol. 20, no. 5, pp. 663–677, 2012.

[3] B. Peerdeman, D. Boere, H. Witteveen, R. Huis in ’tVeld, H. Hermens, S. Stramigioli, H. Rietman, P. Veltink, and S. Misra, “Myoelectric forearm prostheses: State of the art from a user-centered perspective,” *The Journal of Rehabilitation Research and Development*, vol. 48, no. 6, pp. 719–738, 2011.

[4] E. Scheme and K. Englehart, “Electromyogram pattern recognition for control of powered upper-limb prostheses: State of the art and challenges for clinical use,” *The Journal of Rehabilitation Research and Development*, vol. 48, no. 6, pp. 643–660, 2011.

[5] K. Englehart, B. Hudgins, and P. a. Parker, “A wavelet-based continuous classification scheme for multifunction myoelectric control.” *IEEE transactions on bio-medical engineering*, vol. 48, no. 3, pp. 302–11, Mar. 2001.

[6] K. Englehart and B. Hudgins, “A robust, real-time control scheme for multifunction myoelectric control.” *IEEE transactions on bio-medical engineering*, vol. 50, no. 7, pp. 848–854, Jul. 2003.

[7] M. Asghari Oskoei and H. Hu, “Myoelectric control systems-A survey,” *Biomedical Signal Processing and Control*, vol. 2, no. 4, pp. 275–294, Oct. 2007.

[8] D. Bergantz and H. Barad, “NEURAL NETWORK CONTROL OF CYBERNETIC LIMB PROSTHESES,” in *Engineering in Medicine and Biology Society, 1988. Proceedings of the Annual International Conference of the IEEE*, 1988, pp. 1486–1487.

[9] M. A. Oskoei and H. Hu, “Support vector machine-based classification scheme for myoelectric control applied to upper limb.” *IEEE transactions on bio-medical engineering*, vol. 55, no. 8, pp. 1956–65, Aug. 2008.

[10] Y. Huang, K. B. Englehart, B. Hudgins, and A. D. C. Chan, “A Gaussian Mixture Model Based Classification Scheme for Myoelectric Control of Powered Upper Limb Prostheses,” *IEEE transactions on bio-medical engineering*, vol. 52, no. 11, pp. 1801–1811, 2005.

[11] A. D. C. Chan and K. B. Englehart, “Continuous Myoelectric Control for Powered Prostheses Using Hidden Markov Models,” *Biomedical Engineering, IEEE Transactions on*, vol. 52, no. 1, pp. 121–124, 2005.

[12] A. B. Ajiboye and R. F. F. Weir, “A heuristic fuzzy logic approach to EMG pattern recognition for multifunctional prosthesis control.” *IEEE transactions on neural systems and rehabilitation engineering : a publication of the IEEE Engineering in Medicine and Biology Society*, vol. 13, no. 3, pp. 280–91, Sep. 2005.

[13] A. Phinyomark, F. Quaine, S. Charbonnier, C. Serviere, F. Tarpin-Bernard, and Y. Laurillau, “EMG feature evaluation for improving myoelectric pattern recognition robustness,” *Expert Systems with Applications*, vol. 40, no. 12, pp. 4832–4840, Sep. 2013.

[14] D. Farina, A. Holobar, R. Merletti, and R. M. Enoka, "Decoding the neural drive to muscles from the surface electromyogram." *Clinical neurophysiology : official journal of the International Federation of Clinical Neurophysiology*, vol. 121, no. 10, pp. 1616–23, Oct. 2010.

[15] E. Scheme and K. Englehart, "Training Strategies for Mitigating the Effect of Proportional Control on Classification in Pattern Recognition Based Myoelectric Control," *Journal of prosthetics and orthotics*, vol. 25, no. 2, pp. 76–83, 2013.

[16] A. H. Al-Timemy, G. Bugmann, J. Escudero, and N. Outram, "A preliminary investigation of the effect of force variation for myoelectric control of hand prosthesis." in *Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference*, vol. 2013, Jul. 2013, pp. 5758–61.

[17] P. Parker, K. Englehart, and B. Hudgins, "Myoelectric signal processing for control of powered limb prostheses." *Journal of electromyography and kinesiology : official journal of the International Society of Electrophysiological Kinesiology*, vol. 16, no. 6, pp. 541–548, Dec. 2006.

[18] M. Ortiz-Catalan, R. Brå nemark, and B. Hå kansson, "BioPatRec: A modular research platform for the control of artificial limbs based on pattern recognition algorithms." *Source code for biology and medicine*, vol. 8, no. 1, p. 11, Jan. 2013.

[19] M. Ortiz-catalan, B. Hå kansson, and R. Brå nemark, "Real-time and Simultaneous Control of Artificial Limb Based on Pattern Recognition Algorithms."

[20] M. Ortiz-Catalan, R. Brå nemark, and B. Hå kansson, "BIOLOGICALLY INSPIRED ALGORITHMS APPLIED TO PROSTHETIC CONTROL," in *Proceedings of the IASTED International Conference Biomedical Engineering*, no. BioMed, Innsbruck, Austria, 2012, pp. 7–15.

[21] B. A. Lock, K. Englehart, and B. Hudgins, "REAL-TIME MYOELECTRIC CONTROL IN A VIRTUAL ENVIRONMENT TO RELATE USABILITY VS . ACCURACY," in *Proceedings of the 2005 MyoElectric Controls/Powered Prosthetics Symposium*, 2005.

[22] J. Demšar, "Statistical Comparisons of Classifiers over Multiple Data Sets," *Journal of Machine Learning Research*, vol. 7, pp. 1–30, 2006.