Relationships Between Motor Unit Anatomical Characteristics and Motor Unit Potential Statistics in Healthy Muscles

by

Mahdieh Sadat Emrani

A thesis
presented to the University of Waterloo
in fulfilment of the
thesis requirement for the degree of
Master of Applied Science
in
Systems Design Engineering

Waterloo, Ontario, Canada, 2005

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Mahdieh Sadat Emrani

Abstract

The main goal of this thesis was to discover the relationships between MU characteristics and MUP features. To reach this goal, several features explaining the anatomical structure of the muscle were introduced. Additionally, features representing specific properties of the EMG signal detected from that muscle, were defined. Since information regarding the underlying anatomy was not available from real data, a physiologically based muscle model was used to extract the required features. This muscle model stands out from others, by providing similar acquisition schemes as the ones utilized by physicians in real clinical settings and by modelling the interactions among different volume conductor factors and the collection of MUs in the muscle in a realistic way. Having the features ready, several relationship discovery techniques were used, to reveal relationships between MU features and MUP features. To interpret the results obtained from the correlation analysis and pattern discovery techniques properly, several algorithms and new statistics were defined. The results obtained from correlation analysis and pattern discovery technique were similar to each other, and suggested that to maximize the inter-relationships between MUP features and MU features, MUPs could be filtered based on their slope values, specifically MUPs with slopes lower than 0.6 v/s could be excluded. Additionally PDT results showed that high slope MUPs were not as informative about the underlying MU and could be excluded to maximize the relationships between MUP features and MU characteristics. Certain MUP features were determined to be highly related to certain MU characteristics. MUP *area* and *duration* were shown to be the best representative feature for the MU size and *average fiber density*, respectively. For the distribution of fiber diameter in the MU, *duration* and *number of turns* were determined to reflect *mean fiber diameter* and *stdv of fiber diameter* the best, correspondingly.

Acknowledgment

Whoever doesn't thank others, hasn't indeed thanked God.

Apostle of God, Mohammad (pbuh)

This is an opportunity for me to express gratitude to those people who supported me in one way or another. The first person I would like to thank deeply is my supervisor, Dr. Daniel William Stashuk, from whom I have learned a great deal. His breadth and depth of knowledge, hardworking spirit, and experience has always been an inspiration for me. His constant support and guidance led my research toward the right direction and helped me overcome the obstacles.

I would also like to thank Dr. George Freeman and Dr. Otman Basir for spending their precious time on reviewing my thesis and for their comments.

My profound gratitude and love goes to my husband. I would like to thank him for all the joy and happiness he brought into my life and for his great patience, especially in the phases of my intense work on this thesis. I cannot imagine how I could finish this work without his support. Therefore I am dedicating this thesis with all my love and respect to him.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Any body movement, as fast as eye blinking and as slow as intestine contractions, as voluntary as running and as involuntary as breathing, is provided by a series of contractions and relaxations of particular muscles in the body. Healthy muscles convert chemical energy into mechanical form to facilitate motion. However, structural perturbations to muscle anatomy due to different disease processes change functionality and cause particular deficiencies in this tissue. Before treatment, the muscular anatomical changes caused by different disorders need to be discovered. Obviously, good understanding of muscle behavior related to different neuro-muscular disorders facilitates prevention as well.

Physicians tend to study underlying muscle behavior by performing different clinical examinations and tests. The one studied in this project is to analyze the electromyographic (EMG) signals recorded from muscle. An EMG signal is a non-linear, non-stationary potential detected from muscles using electrodes during activation. This complex signal reflects the physiological and morphological aspects of the muscle from which it was recorded. The EMG signal acquired from a diseased muscle is different from that of a normal muscle, since the anatomical structure of the muscle is perturbed due to the disease processes (see Figure 1.1). Hence, discovering the deviations in EMG signals facilitates the understanding of possible alterations in the structure of an underlying muscle.

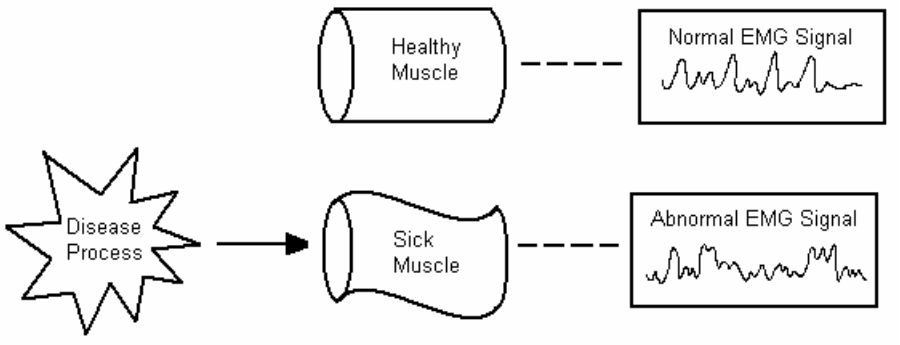


Figure 1.1: Disease processes affect the structure of a muscle and hence alter the shape of the detected EMG signal.

In this thesis, possible changes in EMG signals are studied by discovering different relationships between muscles and the EMG signals detected from them. This is done by studying the associations between structural components of a muscle, known as motor units, and electrophysiological signals detected from them, called motor unit potentials (MUPs). Since the required information regarding the underlying anatomy is not directly available in clinical settings, a physiologically based muscle model is used to generate equivalent simulated data, instead of real data. One main advantage of this model over other simulation models (Nandedkar, 1988a, 1988b, 1988c,Stalberg and Karlsson [2001b,a]), is that the generated EMG signal is acquired based on a specific needle configuration in a clinically valid way. Moreover, detailed parameters affecting the EMG signal are considered and simulated. Hence, the simulated EMG signals are ensured to resemble real EMG signals significantly. These advantages separate the attempt presented in this thesis from similar studies, and facilitate the problem to be investigated from a novel point of view.

From the generated EMG signals, appropriate measurements are calculated to characterize the MUPs. As well, specific features are extracted from the muscle, to represent the underlying physical structure. These features are then used to discover inter-relationships between muscle anatomy and electrophysiology. In order to reveal possible associations

between the two feature types, two different relationship discovery techniques are applied.

Organization of the Thesis

There are five chapters in this thesis including this introduction.

To give a better understanding of the research field, a touch of the background physiology along with a brief review of existing ideas relevant to relationship discovery between muscle physiology and electrophysiology is presented in Chapter 2.

In Chapter 3 a physiologically-based complex muscle model providing simulated clinical EMG signal is described in detail. Also, two relationship discovery methods which are used in the thesis to discover associations existing between EMG signals and the underlying anatomy are explained.

Chapter 4 describes the experimental procedures preceding the data generation routines along with the final results revealing the significant patterns inherent in the dataset.

Corresponding to the final results, a brief discussion is presented in Chapter 5.

Finally, Chapter 6 highlights important contributions of this study and suggests directions for future research in this area.

Chapter 2

Background and Literature Review

The anatomy of a healthy muscle will be damaged if affected by a disease process. As a result of alteration in its physical properties, the muscle's functionality perturbs as well. One way of investigating the alterations due to disease is to analyze EMG signals detected from the diseased tissue. To do this, a good understanding of the relationship between EMG signals and the underlying muscle anatomy is needed. In this chapter some morphological and physiological aspects of muscle are introduced. In addition, some electrophysiological features of muscle and detected EMG signals, along with associated information presented in the literature are mentioned and briefly discussed.

2.1 Muscle Physiology

We use our muscles to convert chemical energy into mechanical work. Mechanical work or movement is the result of contraction of muscles. There are three kinds of muscles found in our body.

Heart Muscle or **cardiac muscle** which makes up the wall of the heart. Throughout life, it contracts some 70 times per minute pumping about 5 liters of blood each minute.

Smooth Muscle which is found in the walls of all hollow organs of the body (except the heart) and its contraction reduces the size of these structures. Thus it regulates the flow of different kinds of liquid material associated with hollow organs, such as blood in arteries, food in the gastrointestinal tract and so on.

Skeletal Muscle which is connected at either or both ends to a bone and moves parts of the skeleton. It is used to facilitate limb movements as well as to maintain different body postures.

In this project physiological and morphological features of **skeletal muscle** along with electrophysiological features of electromyographic (EMG) signals, were investigated. Consequently inter-relationships between features of these two sets were revealed. From now on, the term *muscle(s)* will refer to *skeletal muscle(s)*.

The structural unit of contraction is the muscle cell or muscle fiber. Best described as a very thin thread, the muscle fiber has a length ranging from a few millimeters to 30 cm and a diameter of 10 μm to 100 μm. On contraction it will shorten to about 57% of its original resting length (Basmajian and Luca [1985]). Each muscle fiber is supplied by a terminal branch of one nerve fiber or axon whose cell body is located in the spinal cord. However, each motor neuron (MN) innervates[1] a group of several muscle fibers. The muscle fibers innervated by one MN plus the nerve cell body, the long axon running down the motor nerve and its terminal branches, together constitute a motor unit (MU). However, the MU territory is defined typically on the muscle level by the cross-sectional area of the cylindrical volume including all muscle fibers innervated by the associated MN. A MU is the functional unit of skeletal muscle, since an impulse propagated from the MN causes all the muscle fibers in one motor unit to contract almost simultaneously. The anatomical characteristics of a healthy muscle follow specific distributions, e.g. muscle fiber diameters in a MU are Gaussian distributed with a mean of 55 μm and standard deviation of 9 μm typically.

[1] supply the nerve to

There are several disease processes associated with muscles and nerves, known as neuromuscular disorders. Divided into two main categories of myopathy and neuropathy, they perturb the structure and functionality of muscle by affecting muscle fibers (in myopathy) and motor neurons (in neuropathy). Myopathy is mostly characterized by muscle fiber atrophy[2], loss and replacement by connective or fatty tissue, hypertrophy[3] and splitting into several thinner muscle fibers. In Figure 2.1 two sets of muscle fibers are shown in biopsies taken from myopathic cases. Variation in fiber diameter is obviously seen in one set, while fiber splitting is the dominant effect in the other one.

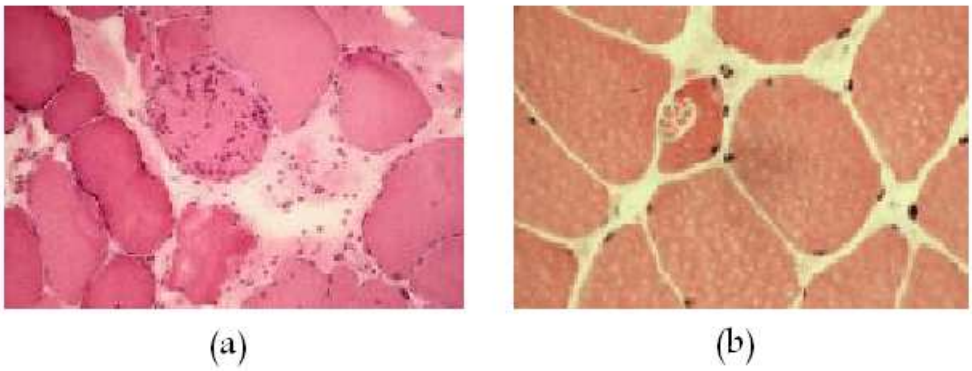


Figure 2.1: Two cross-sectional views of biopsies taken from a specific myopathic tissue (Limb-Girdle Muscular Dystrophy). Note the variation in fiber diameter (a), and the big fiber which has split into several tiny fibers (b).

On the other hand in neuropathies, MN loss is an early sign, where affected fibers, i.e., those associated to dying MNs, lose their nerve supply and get denervated. As a result, unaffected MNs in the immediate vicinity of orphan[4] muscle fibers produce collateral nerve sprouts and re-innervate[5] the denervated muscle fibers.

[2]decrease in size as a result of disease or disuse

[3]abnormal enlargement of a body part

[4]denervated muscle fibers

[5]supply the nerve for

There are different diagnostic methods for providing detection of abnormalities caused by neuromuscular (NM) disorders. Physical examination, muscle biopsy and laboratory examinations (protein synthesis, genetic testing) are the most common ones used (Bethlem [1980]). Another way to recognize the structural changes caused by disease is to study the electrical activities of the muscle quantitatively. This information is obtained from the analysis of EMG signals detected from the muscle. An EMG signal is a complicated composition of the electrical potentials detected from muscle fibers during activation.

2.2 Muscle Activation and Electrophysiology

Contraction of muscle fibers is preceded by initiation of an action potential on the muscle fiber in the endplate region or at the neuromuscular junction[6] (NMJ). The action potential will propagate along the length of the fiber in both directions to each tendinous end. In healthy muscles, an active MN concurrently stimulates all muscle fibers connected to it. The currents associated with these potentials spread throughout the extracellular volume surrounding muscle fibers and create time varying potential fields. Electrodes positioned in the extra-cellular space detect the latter field potentials and the resulting measurements can be expressed as waveforms.

Waveforms generated by single muscle fibers are called muscle fiber potentials (MFPs) whereas those associated with a motor unit are called motor unit potentials (MUPs). To maintain or increase the force generated by a muscle, MUs become activated frequently, hence creating a train of MUPs associated with each MU. The number of recurring MUPs along a train defines the MU firing frequency. The collection of the motor unit potential trains (MUPTs) detected during a contraction is termed an EMG signal. Basically an EMG signal is the combination of motor unit potentials (MUPs), each of which is a composite of muscle fiber potentials (MFPs) generated by the muscle fibers of the MU. It

[6] the connection between nerves and muscles

is an exceedingly complicated signal which is affected by the anatomical and physiological properties of the muscle, the control scheme of the peripheral nervous system, as well as the characteristics of the instrumentation that is used to detect and observe it.

However, it is the morphological and physiological features of the active motor units and their muscle fibers, which form the basis for clinical EMG signals. An EMG signal contains valuable information about the underlying muscle from which it is detected. There are three main approaches in conventional clinical electromyography for using EMG signals to assess the structure and activation pattern of a muscle (Stashuk and Brown [2002]). The first approach is to examine the spontaneous electrical activity of the muscle at rest, for which suitable quantitative techniques have not yet been developed. The second approach examines the shapes and sizes of the individual MUPs of voluntarily recruited MUs. This provides important clues regarding the sizes and innervation patterns of the associated MUs and how they may have changed in response to a neuromuscular disorder. In the third approach the composite EMG signal detected during voluntary contraction (the so-called interference pattern (IP)) is analyzed to assess the activity and size of individual MUs. Fortunately, for the two latter cases, quantitative techniques are available.

In most clinical electromyography laboratories, the common practice is to 'eyeball' EMG signals, in other words to analyze the signal in a subjective and qualitative manner. The electromyographer compares the recorded signal against personal databases collected from examining several healthy subjects of the same age and patients with a variety of NM diseases. Obviously, the final diagnosis is biased significantly based on the electromyographer's experience and the content of the databases. On the other hand, quantitative electromyography (QEMG), offers a more objective and semi-automatic analysis by quantifying measurements of EMG signals. In a QEMG study different distributions of selected features of EMG signals acquired from any given subject and muscle are used to suggest whether the underlying muscle is affected by a specific type of NM disorder or not. Based on the nature of a QEMG study, it embraces a vast variety of techniques. The one used in

this work, quantifies the EMG signal by extracting different feature values from individual MUPs. The basic premise behind all of these techniques is that specific features of an EMG signal represent specific physiological and morphological characteristics of the underlying muscle. Pathophysiology can therefore be assessed based on features of detected MUPs. As an indirect approach it requires a detailed knowledge of the relationships between the MUP generators and the MUP measurements. Simply due to the lack of practical methods to identify specific MUs in a muscle, it is difficult to study the inter-relationships between MUP features and specific characteristics of the underlying motor units. Furthermore the co-variation among different types of pathology has made it impossible to extract the influence and importance of each individual feature. To overcome these problems different mathematical models and computer simulations have been developed to facilitate studying the influence of individual morphological parameters of a MU on specific MUP features. Obviously some features represent neuromuscular changes more clearly than others.

Buchthal and co-workers introduced MUP quantitative evaluation in the 1950's and measured amplitude, duration, number of phases and number of turns of detected MUPs. They concluded that various myopathies and neurogenic disorders could be distinguished from one another using these features (Buchthal and Pinelli [1952], Buchthal et al. [1954]). Since then, these parameters have been studied and their relationships with anatomical characteristics of the MU have been analyzed further. Although there are no specific universal definitions for MUP parameters, they have been characterized in the AAEE[7] glossary of terms in clinical electromyography. Amplitude is defined as the voltage difference measured from the most positive peak to the most negative peak. Clinical EMG studies (Takehara et al. [2004], Barkhaus et al. [1990], Dumitru et al. [1997]) have revealed that amplitude is derived primarily from fibers within 300-500 μm of the electrode's recording surface. The greater the number of near fibers or the closer they are to the electrode, the larger the MUP amplitude. As confirmed by simulation studies (Nandedkar et al. [1988a],

[7] American Association of Electromyography and Electrodiagnosis

Nandedkar and Sanders [1989], Nandedkar et al. [1988c], Nandedkar [2002], Stalberg and Karlsson [2001b,a]), MUP amplitude represents the size and number of the closest fibers to the electrode. Thus it is sensitive to needle position; slight changes in needle position will change the MUP amplitude significantly (Barkhaus et al. [1990]). MUP duration on the other hand is less sensitive to the distance between the needle and active muscle fibers of a MU, since it is determined by a much larger percentage of the muscle fibers of a motor unit (Buchthal et al. [1954]). It is considered a robust diagnostic parameter and is defined as the time interval between the potential's departure from, and subsequent return to the baseline. It depends on the MU muscle fibers that are up to 2.5 mm from the electrode (Nandedkar et al. [1988c], Nandedkar [2002]). Considering a typical MU with a circular territory diameter of 7 mm, the uptake area for duration would include roughly 50% of the muscle fibers in the MU. Therefore duration is essentially related to the number of fibers in the MU or the size of the MU and it is believed that changes in MUP duration reflects changes in MU size (based on simulation models, Stalberg and Karlsson [2001a]). This makes duration useful in differentiating myopathy from neuropathy. However, MUP duration changes with increase in the dispersion of arrival times of MFPs at the electrode; this affect can actually overcome the MU size factor (Nandedkar and Sanders [1989]). Long-duration complex MUPs associated with MUs that have lost muscle fibers have been found in chronic myopathies (Uncini et al. [1990]). Therefore it is suggested to exclude complicated MUPs for MU size assessments using duration measurements (Dumitru et al. [1999], Dumitru and King [1999], Uncini et al. [1990], Barkhaus et al. [1990]). Simulation studies revealed that duration measured only from simple MUPs reflects the size of the associated MU (Nandedkar and Sanders [1989]). Another parameter used in quantitative MUP studies introduced by Buchthal, is phase, defined as the portion of a wave between the departure from and the return to, the baseline. A turn is described as the point of change in slope of the waveform and the magnitude of the voltage change following the

turning point[8]. Together number of turns and phases, measure the complexity of the signal (Buchthal et al. [1954]), often increased in pathological muscles, caused by diameter variation of the affected muscle fibers (Stalberg and Karlsson [2001a], Nandedkar et al. [1988b]) within 1 mm of the electrode core (Nandedkar [2002]). It was believed that fiber loss relates to polyphasic MUPs (MUPs with more than 4 phases) in myopathies, but simulation studies described the variability in fiber diameter as the only cause (Nandedkar and Sanders [1989]). Clinical studies (Pfeiffer and Kunze [1992]) revealed that phase counts provide no additional information in discriminating between normal and myopathic MUPs. However, they are complementary to turn counts in discriminating between normal and neurogenic muscle. To further quantitate MUP complexity, the irregularity coefficient was introduced (Zalewska and Hausmanowa-Petrusewicz [1995]) through which the normalized length of the MUP curve was measured providing a convincing link with a visual estimate of the signal irregularity. Since this coefficient characterizes the detailed features of a potential (Zalewska and Hausmanowa-Petrusewicz [1999]), it evaluates MUPs with respect to disease progression regardless of the myogenic or neurogenic origin of the disorder (Zalewska and Hausmanowa-Petrusewicz [2000]).

Another feature used for evaluating a MUP quantitatively, is the area under the rectified potential curve. It depends on muscle fibers that are up to 2 mm from the electrode tip and is reduced with fiber loss and increases when fibers are regenerated (Nandedkar and Sanders [1989]). The are is determined by two main factors: the number and the diameters of muscle fibers (amplitude) and the temporal dispersion of their MFPs (duration) (Stalberg and Karlsson [2001b], Cengiz et al. [2002], Brooke and Engel [1969b]). Being related to both amplitude and duration it would be affected by needle movement, although it is less sensitive compared to amplitude (Takehara et al. [2004]). Normalization with amplitude reduces the influence of needle position. Therefore the $area/amplitude$ ratio or thickness

[8]It is not necessary that the voltage change passes through the baseline. A minimal excursion required to constitute a change should be specified.

was introduced (Nandedkar et al. [1988a]) to further quantify MUP shapes. It was shown that this ratio is useful in discriminating between myopathic and normal muscles. However, it wasn't better than other MUP parameters in separating neurogenic from normal muscles. Sonoo and Stalberg (Sonoo and Stalberg) used discriminant analysis to develop a simplified function including both amplitude and thickness:

$$2 \times log_{10}(amp_{mV}) + (area/amplitude)_{ms} \tag{2.1}$$

called the size index (SI). Clinical studies have indicated that SI is a strong discriminative parameter between neurogenic and normal muscles (Zalewska and Hausmanowa-Petrusewicz [2000], Sonoo [2002]) and a stable parameter (Sonoo and Stalberg, Takehara et al. [2004]) that is not influenced by needle position. Moreover, it is the most reproducible parameter since it is least dependant on technical and patient factors and it is the best parameter to follow in sequential QEMG studies (Takehara et al. [2004]).

In addition to the features mentioned above, maximum slope, maximum acceleration and fiber count are also considered as useful MUP features. Maximum slope is the peak of the slope, the first derivative of MUP waveform, and is not used as a diagnostic feature, but rather as a criterion or index of utility of detected MUPs (Barkhaus and Nandedkar [1996]). In clinical electromyographic recordings only MUPs containing sharp spikes are used for assessment, while sharpness is measured through MUP's maximum slope. Besides, it is believed that maximum slope associates with the electrode relative distance to MU, the influential factor but yet unknown in clinical EMG acquisition. Correspondingly, maximum slope is suggested as a useful parameter being used in this project, describing the status of MU regarding the location of electrode tip. Another feature also considered as an index throughout this work, is the maximum acceleration or the peak of second derivative of MUP waveform, and is used in detection of single fiber contributions to motor unit potentials. Simulation studies have shown that the activity of individual muscle fibers contributing significant MFPs to MUPs of a MU can be studied through analyzing MUP accelerations (Stashuk [1999b]). Fiber count is measured by counting the number of sufficient sharp

acceleration peaks with amplitude above some threshold and is chosen in this project as a proper MUP feature. It represents the number of fibers in close proximity of the electrode (Stashuk [2001]).

Thus far, MUP features commonly used to interpret the morphology[9] of the underlying active MUs have been introduced. As well, the status of an underlying MU can be characterized by a number of morphometric[10] features. These features reflect the alterations in anatomy due to NM disorders.

Since an increase in fiber diameter variation is one of the early signs of NM disease (Wang et al. [1999], Schoser et al. [2004]), mean fiber diameter and fiber diameter variations were selected as MU parameters. For the first time, in 1969 Brooke and Engel analyzed and measured muscle fiber diameters in 103 normal biopsies (Brooke and Engel [1969a]). They constructed histograms including mean diameter, standard deviation and the number of fibers measured. Moreover, equivalent experiments for 182 biopsies of neurogenic (Brooke and Engel [1969b]) and 97 myogenic (Brooke and Engel [1969c]) cases were performed and the proportions of abnormally large and small fibers were indicated. Semi-quantitative morphometric techniques were used to determine the distribution of fiber diameter of biopsies performed in healthy cases (Doriguzzi et al. [1984]) and patients affected by various NM disorders (Johnson et al. [1973], Cros et al. [1989], Tohgi et al. [1994]). Increase in the coefficient of variation (cov) was shown to be more frequent and pronounced in neurogenic processes than in myoapthic syndromes (Tosi and Jerusalem [1976]), in which fiber enlargement or hypertrophy was reported to be dominant compared to atrophy (Verma et al. [1992]). Fiber diameter is believed to be the most sensitive feature to disease processes, characterizing muscle morphology.

Fiber density (FD) or number of fibers per mm^2 of MU territory is another MU parameter used, alteration of which is mentioned as the earliest sign of denervation seen

[9] branch of biology that deals with the structure of animals and plants

[10] related to morphometry(method that involves measurement of the shape)

in neuropathies (Stalberg and Karlsson [2001a]) and some myopathies (Bertorini et al. [1994]) (Wang, 1999). In addition, the number of fibers per MU is used as a representative parameter for MU size (Nandedkar et al. [1988c]).

As mentioned earlier, during clinical studies, other factors beside MU anatomy and physiology such as the type, position and rotation of the electrode, intracellular tissue conductivity, etc. affect the detected MUPs. These so called volume conductor factors (VCF) are neither related to MU morphology nor extractable from MUPs, but they do alter MUP shape and size. Among all volume conductor factors the greatest impact on MUPs belongs to the electrode through which the EMG signal is detected. The electrodes used in electromyography are of a wide variety of types and construction. However, there are two main types: surface (skin) electrodes and inserted (wire or needle) electrodes. Each has its advantages and limitations. Having a relative large pickup area, surface electrodes are commonly used to detect gross EMG signals consisting of the electrical activity of numerous individual MUs, through which valuable information regarding the sizes of MUs can be obtained. However, surface electrodes can't be used to detect signals selectively from small groups of fibers due to cross-talk effects of adjacent muscle fiber groups. Therefore detailed anatomical features of the muscle can not be obtained using surface electrode acquired signals, which makes it impractical as a choice of this work.

On the other hand, the small pick up area of a needle electrode enables detection of individual MUPs conveniently, permitting a fine and selective study of each MUP. Thus it is possible to study features of individual MUP and investigate their relationship with specific MU characteristics. Another advantage of needle electrodes over surface electrodes is that it may be easily repositioned within the muscle so that new territories can be studied or signal quality may be improved (Basmajian and Luca [1985]). The most common needle electrode is the concentric needle (CN), consisting of an elliptical tip[11] and a relatively

[11] with a minor and major axes of typically 150 μm and 580 μm.

long cannula[12] or reference detection surface. It is widely used for clinical and research applications. Beside the type of electrode, its position and rotation affect the shapes and sizes of detected MUPs. Nandedkar *et al.* (Nandedkar et al. [1988c]) performed correlation analysis between electrode position and some MUP features. They defined electrode position as the electrode distance to the nearest fiber of the MU, with which the correlation coefficient of MUP amplitude was found to be -0.77, suggesting that the greater the distance to the MU fiber the smaller its contribution to MUP amplitude. The described distance depends on knowing the MU territory which is not the case in actual clinical recordings therefore simulation models have been used to study the effect of distance on different MUP features. However, real clinical recordings are available regarding the electrode distance to neuromuscular junction region; Falck et al. (Barkhaus et al. [1990]) have studied MUPs detected from different recording sites in brachial biceps[13] by positioning the needle in the middle third and distal third of the muscle belly. They have shown that the more distant the needle is from the endplate, the more pronounced is a MUP's temporal dispersion and MUPs with longer duration are detected.

The goal of this work is to find possible relationships between anatomy and electromyography by investigating the inter-relationships between specific MU characteristics and MUP features. To achieve this, relationship discovery techniques are applied to MU and MUP features characterizing anatomical and electrophysiological aspects of the muscle. In this regard, others have performed correlation analysis between different MUP features for real[14] clinical data in normal and diseased cases (Zalewska and Hausmanowa-Petrusewicz [1999, 2000], Cengiz et al. [2002], Pattichis and Elia [1999]). However, to reveal the associations between different MUP features and MU characteristics, simulated models have been used (Nandedkar et al. [1988a,c,b], Stalberg and Karlsson [2001b,a]), simply because precise MU structure is unknown during clinical EMG signal acquisition. Similarly in this

[12] 25 mm - 75 mm

[13] the muscle that flexes and supinates the forearm

[14] versus simulated data

project simulated data was used to assess the relationships mapping specific anatomical characteristics to specific electrophysiological features.

There are many different simulation models used to assist in MUP quantitative evaluation. Some use pure mathematical representations which focus on the overall statistical and spectral properties of the detected EMG signals, while others combine a form of the MUP shape along with simulated MU firing times to emulate an EMG signal (Stashuk [1993]). There are different methods used in various models to imitate the MUP shapes. Arbitrary analytical estimation for the shape of the MUP is one method, whereas MUPs extracted from real EMG signals are used as another method to substitute the MUP in simulation models. In this work, a muscle model was used to explore the relationships between changes in muscle structure and values of characteristic features extracted from MUPs. Therefore the simulated EMG signals must resemble real EMG signals and reflect the contribution of MU architecture to the values of specific MUP features. To achieve this, individual MUP shapes are constructed based on specific electrode and MU configurations. Basically a primary structure representing the underlying anatomy and morphology of a muscle is built and then a specific electrode configuration is considered to detect the simulated potentials generated from activated muscle fibers representing the physiology and morphology. As for other methods, the MUP waveforms are combined along with simulated firing times to build up a composite EMG signal. Many researchers have utilized this MUP generation methodology and used a muscle model as the basis for their EMG simulation. Nandedkar *et al.* (Nandedkar et al. [1988c]) have used such a model for normal cases in finding MUP feature relationships. For abnormal situations, either myogenic (Nandedkar and Sanders [1989]) or neurogenic (Nandedkar et al. [1988b]) they have perturbed the normal model of MU architecture appropriately to be able to investigate the individual and combined effects of these processes on the shapes of MUPs. A similar model has been used by Dumitru *et al.* (Dumitru et al. [1999], Dumitru and King [1999]) to study MUP duration and investigate the effects of muscular structure on this parameter in more detail.

Stalberg and Karlsson (Stalberg and Karlsson [2001b,a]) in an effort to simulate concentric needle electromyography have utilized a muscle model in their MUP generation scheme. They concluded that the simulation model indicates the relative sensitivity of various EMG parameters and gives the opportunity to study any MU characteristic quantitatively.

The model used in this work embraces the same methodology for MUP generation. The rationale leading to the formation of the general model was originally introduced by Stashuk (Stashuk [1993]). This method is based on the physiology and structure of muscle and begins with modeling a cross-section of a muscle. Within this cross-sectional area, MU centers are randomly located. Then muscle fibers are randomly assigned to MUs based on their territory sizes. For the MUs with fibers in the detection area, firing rate behaviors as a function of an assumed level of contraction are then simulated. Motor unit potentials for a specific needle configuration are created and finally the detected MUPs are combined with respect to the MU firing patterns to create the composite EMG signal. This model has been evaluated qualitatively by comparing the results to either real signal properties or those of other models in the literature (Stashuk [1993]). Different aspects of muscle morphology and physiology are simulated in this model.

Recently, this model has been extended to include developed algorithms that generate EMG signals consistent with those acquired in a clinical setting (Hamilton-Wright and Stashuk [2005], Hamilton-Wright et al. [2002]). The extended model includes certain enhancements compared to other models in the literature.

In the current model a typical muscle including 150 - 250 MUs is simulated in each experiment, compared to previous models where just one simple MU at a time was simulated. The process of acquiring an EMG signal is constructed in a similar way to the one utilized in present clinical settings where the physician searches for an active fiber by moving the needle after insertion, before starting the recording. While in most models the needle is positioned randomly relative to the active fibers. NM disorders cause various physiological changes preceding anatomical alterations in neurons and muscles. However, in most

models the changes are studied one by one for simplification reasons. In this model all known anatomical perturbations are simulated concurrently, and their accumulated effects on MUP features are studied. One may argue that simplified models, in which some restricted parameters are simulated, are the first steps in understanding the functionality of a complex system, and that a complicated model considering cumulative effects of different parameters is difficult to understand and study. However, we should keep in mind that different changes to MU anatomy due to a NM disease process occur together affecting the detected signal simultaneously. Therefore it is necessary to study the additive effects of different anatomical alterations on different MUP features together, which is possible using the described model.

The significant advantages over other models presented in the literature, motivated us to conduct similar studies using this uniquely different model. Considering the interactions existing among groups of active MUs along with various morphological parameters concurrently, highlights this simulation model among others. Moreover, including an EMG signal acquisition scheme similar to the one used by physicians in clinical settings, increases the chances that the detected EMG signals will resemble real clinically detected EMG signals. Therefore it is well justified to use this model as a reliable source for discovering associations between real clinical MUPs and MUs obtained from real muscle. Besides, it is interesting to seek the same relationships between MUPs and MU features extracted using this complex model, as the ones mentioned by others using more restricted models.

In this model the developed algorithms determine a muscle model in advance and allow independent positioning of an electrode and selection of contraction level. Therefore, the structure and activation of a simulated muscle can be sampled at different needle positions and levels of contraction as would be done during acquisition of real clinical EMG data. This muscle model along with its EMG signal simulation algorithms have been shown to produce EMG signals consistent with those acquired from real muscle (Hamilton-Wright and Stashuk [2005]). Therefore it creates the unique opportunity to deal with semi-real

data, compared to other simulation models, and yet have sufficient information regarding the underlying anatomy, compared to actual real data.

Since the anatomical structure of the model resembles real muscle and the detected EMG signal provides clinically relevant details, the data from such a model includes valuable information about the properties of real muscles and their relevance with EMG signals that would be detected during a clinical study. Hence, relationships detected would expect to exist between real muscles and detected EMG signals. Therefore the extracted MU characteristics and MUP features are reliable enough to be used for studying inter-relationships between underlying anatomy and acquired EMG signals. This muscle model along with the developed algorithms for EMG signal generation, are described in detail in the next chapter of the thesis.

Having extracted the appropriate features, relationship discovery techniques should be used to find the associations between EMG signals and underlying muscle anatomy. Correlation analysis was performed as the most common method used in discovering linear associations in a dataset. Additionally a pattern discovery technique was used to find association patterns inherent among various features and their values. There are specific advantages for each of these association discovery methods which are discussed in the next chapter.

An obvious prerequisite of the study of any pathological alteration in neuromuscular tissues is to obtain the best possible estimate of normality in healthy cases. Considering the normal values, the deviations are then recognized and measured as signs of pathology and extent of progression. Therefore normality measurements are always stressed to avoid false positivities in diagnostic evaluation of EMG signals with minimal changes. That's why the study in this work has focused on normal data acquired from healthy tissue.

Chapter 3

Data Generation and Analysis Methods

An electromyographic (EMG) signal is the voltage detected from a muscle during activation using suitable electrodes. It contains valuable information regarding the architecture and functionality of the muscle from which it was detected. Thus it reflects changes in muscle morphology and physiology due to disease processes.

To properly interpret the information inherent in an EMG signal associated with the status of the muscle from which it was detected, relationships between anatomical characteristics of muscle and features of a detected EMG signal were studied. To do this, appropriate MU and MUP data were obtained using a complex electrophysiological muscle model. Having the data, several relationship discovery techniques were then applied, in order to find existing-yet-unknown associations among the attributes of the data. The analysis techniques along with the anatomic model and the EMG signal generation scheme used in the thesis, are presented and discussed in detail in this chapter.

3.1 EMG Data Generation Scheme

During clinical electromyography study, physicians try to understand muscle behavior by assessing detected EMG signals. However, many influential anatomical factors are unknown to them at the time. Therefore, it is impossible to directly relate an acquired EMG signal to the morphology and physiology of the underlying muscle in real clinical settings. In an effort to repair this lack of knowledge, researchers have suggested various ways of simulating EMG signals. In this project a detailed physiologically-based simulation model is used to study the structure and activation of a muscle using components of acquired EMG signal. Analysis of EMG data extracted from such a model provides valuable information regarding real muscle structure and its correlation with detected EMG signals. Through this model the opportunity to simulate various pathological situations is obtained.

In this model, an EMG signal is simulated based on a specific muscle structure and needle configuration by rigorous determination of individual MUP shapes along with simulated MU firing patterns. Therefore, the simulated signals resemble real EMG signals at all levels of analysis, from general shape characteristics, to subtle features regarding contributions of individual muscle fibers. Moreover, the EMG data generated by this model should reflect the structure and activity of a muscle in a way similar to the reflection provided by real signals detected from real muscle. Important components of this muscle model have been described and evaluated elsewhere (Hamilton-Wright and Stashuk [2005]) to determine the ways and extent to which the new simulation model can be used. In this section, the specific structural and functional model of the muscle along with the needle electrode configuration is described.

3.1.1 Muscle Structure Model

The method of EMG signal simulation utilized in this project, embraces constructing an anatomical model for the muscle as the base. The physical layout of normal muscle is

formed through several steps which are mentioned in detail elsewhere (Stashuk [1993], Hamilton-Wright and Stashuk [2005]) and briefly in the following.

The morphology of the motor unit (MU) as the building block of the muscle is constructed first, by calculating the MU territory diameter regarding its twitch[1] tension. Using this diameter as the expected MU size along with MU muscle fiber density and area per fiber (0.0025 mm^2), the area of the muscle is estimated. The muscle radius is then calculated and the position of MU centers are determined. Just like seed scattering across a surface, MU centers are 'thrown' uniformly from grid points across the muscle cross sectional area. In order to avoid clustering, grid points are chosen randomly and without replacement until all of them are used. Then all points are reselected for subsequent 'throws' until all MU centers are positioned.

Once MU centers are located, fibers are assigned to MUs in two steps. First fibers are placed based on a fixed grid assuming 400 fibers in each mm squared. Then in the second step each fiber is assigned to a specific MU based on several developmental factors each of which indicates the plausibility of the muscle fiber belonging to the MU. A weighted combination of these factors is then considered to determine the likelihood of a specific MU adopting a specific MF.

Once 20 fibers are adopted by a given MU, the MU center is relocated to the center of mass of the 20 fibers as a whole set. This recalculation is done in order to decrease the number of outliers. The MU center is updated for each muscle fiber adopted thereafter. After all fibers are assigned to different MUs, the actual MU territory diameters are calculated based on the furthest 10% of fibers in each MU.

The fiber diameters of each MU are selected from a Gaussian distribution specific to each MU such that the collection of fibers diameters from all MUs will have a distribution similar to a target overall muscle Gaussian distribution with mean of 55 μm and standard deviation of 9 μm. Considering the difference in size of type I and type II fibers and

[1] response of the muscle to a single stimulus

the predominance of type I fibers in smaller MUs, mean fiber diameter for each MU is expected to not be the same and directly related to MU territory size. Therefore, the mean fiber diameter for each MU is assigned based on the number of muscle fibers in that MU, from a range of fiber diameters, considered equal to 2 times the standard deviation of fiber diameters in the muscle. A special scheme is used to do this assignment such that the sum of MU fiber diameter distributions resembles the overall-muscle distribution of fiber diameters. This scheme lets each MU have less scattered fiber diameters than the whole muscle has and lets larger MUs have both larger fibers and larger variation of fiber diameters.

Last step in physical construction is to determine the end-plate position or neuromuscular junction (NMJ) location for each fiber. As for the distribution of fiber diameters, the distribution of NMJ locations for a MU is less deviated compared to the whole muscle and is assumed to be Gaussian distributed with a mean of 0 μm and a standard deviation based on the MU territory diameter.

So far the muscle architecture is built and the next stage is to simulate the functionality of the muscle during activation. To acquire an EMG signal during activation, motor unit potentials created by active muscle fibers in the MUs, are detected by a needle electrode. A specific electrode configuration is constructed to substitute the real needle electrode in this model and it is discussed prior to signal simulation.

3.1.2 EMG Needle Electrode and Clinical EMG Studies

Several types of needle electrodes are used for clinical EMG studies. Each of these electrode configurations consist of a relatively large cannula or distant reference detection surface and a relatively small tip or side-port detection surface, the size of which varies for different needle structures. The most common one is the concentric needle electrode (CNE) and is used in the model presented here.

In the simulation model the position of the electrode tip is typically specified, to explore

various parts of a muscle. However, in a clinical EMG study, a physician is additionally interested in positioning the needle tip next to active fibers to acquire a 'sharp' signal. To achieve this, the physician searches a small volume of the muscle by rotating or shifting the cannula and moving the tip. Therefore, in order to properly model clinical studies, the searching activity of the physician is emulated by refining the electrode position to ensure detection of a 'sharp' signal, described as a signal containing sufficiently short rise-time (< 500 μs) and adequate amplitude (>50 μv). To do this, the needle tip position is repositioned from the specified needle location to 'nearby' active fibers by moving the tip within an 800 μm radius. A 'nearby' fiber is defined as a big close fiber and is quantified by a large value of *fiber radius/distance* ratio, where '*distance*' is the fiber's distance from the center of the current electrode detection surface location. The tip is positioned at the centroid of 3 'nearby' active fibers.

After inserting the needle, in real muscle tissue, the fibers beside the electrode are repositioned by the electrode mass. To simulate this, when the tip is positioned, the fibers intersecting the needle cannula are displaced from their original locations, depending on their position relative to the needle by moving them to the nearest boundary of the needle cannula. They are projected perpendicularly to the cannula edge or pushed below the tip, whichever is closer (Hamilton-Wright and Stashuk [2005]).

3.1.3 EMG Signal Generation

Now that the physical structure and needle configuration is described, the voltage potentials acquired from each active muscle fiber can be calculated for a given needle tip and cannula position.

Muscle Fiber Potential Modelling

As mentioned earlier a motor unit potential, the unit forming the complete EMG signal, is the superposition of spatially and temporally dispersed muscle fiber potentials detected

during activation. Therefore, muscle fiber potential computation is a convenient starting point for EMG signal modelling. To do this a line source model, based on convolution of a current signal and a weight function (Stashuk [1993], Hamilton-Wright and Stashuk [2005]) is used. The current signal represents the impulse initiated at the endplate with a magnitude depending on the fiber diameter and internal conductivity. However, the weight function models the propagation of this impulse along the muscle fiber length and considers the location of the fiber relative to the detection surface (either tip or cannula of the needle) for a specific needle configuration. For the large exposed detection surface (major and minor axis of 580 μm and 150 μm) of a concentric needle electrode, the weight function is estimated by averaging over 6 narrow, equally spaced line electrodes (Hamilton-Wright and Stashuk [2005]). However, for the cannula which possesses linear shape rather than planar, the weight function is modelled as a single line electrode. Based on the above explanations, it is reasonable for the convolution of the weight function and the current signal to be the muscle fiber potential acquired at a defined detection location. This method of MFP modelling has been used by many other researchers (Nandedkar et al. [1988a,b,c], Nandedkar and Sanders [1989], Duchene and Hogrel [2000], Stalberg and Karlsson [2001b,a], Nandedkar [2002]). It assumes that muscle fibers are straight, cylindrical and placed in a medium with cylindrical anisotrophy. Fiber length and propagation of the action potential along the fiber both toward and away from the detection surface are considered in this model.

Fiber length assumption

The length of the fiber is considered to be finite in this model, causing the convolution of the current signal (Figure 3.1-a) and weight function (Figure 3.1-b) to produce two artifacts regarding the endplate region and the tendons at both ends of the fiber. To remove these artifacts, they were forced to be zero in a fixed way in the former EMG signal generation algorithm (Stashuk [1993]).

However, to provide a more realistic signal acquisition routine consistent with that of

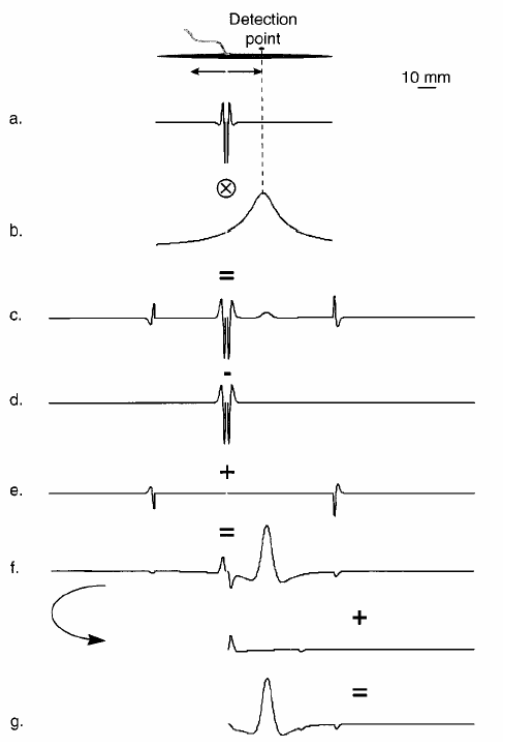


Figure 3.1: Computation of muscle fiber potential. Vertical scale: arbitrary units (dilated vertical scales for f and g plots). (a) Current signal. (b) Weight function. (c) Pure convolution result. (d) Neuromuscular junction effect. (e) Tendon effects. (f) Signals in both directions after end effects compensation. (g) Final potential. (Duchene and Hogrel [2000])

clinical settings, a specific technique for eliminating the artifacts was added as part of the model to the current model. In this technique, which is described by Duchene (Duchene and Hogrel [2000]), compensations regarding the initiation and extinction effects are computed and subtracted (Figure 3.1d-e) from the pure convolution result. The signal propagating away from the detection surface is then flipped and added up to the signal travelling toward the detection surface (Figure 3.1-f). Through these steps the final muscle fiber potential detected by the specific electrode is created.

Motor Unit Potential Modelling

A single motor unit potential (MUP) detected at a specific point results from the summation of the fiber potentials of all the fibers belonging to the same MU.

In this model contribution of an MFP to a specific MUP depends on the fiber's relative location to the electrode which determines the frequency content of the potential created by the fiber. The potential of those fibers far from the electrode, composed of low frequency content are summed up together in just one buffer while MFPs of fibers near the electrode are stored separately. The latter termed as 'near' MFPs , contain high frequency content and are sampled 30 times faster than the former 'distant' MFPs.

MUPs are assumed to be affected by three factors: 1) Fibers in the same MU may have slightly different diameters, hence different conduction velocities, causing a temporal dispersion between individual MFP contributions to a MUP. 2) Chemical factors within each NMJ create variations in the delay between the arrival of the axonal impulses and the initiation of MFPs which in turn varies (causes) the temporal dispersion of individual MFPs (to be different) during each activation. 3) Obviously each MUP is affected by the position of the needle tip detection surface relative to the position of its muscle fibers.

The fixed temporal dispersion, 1^{st} factor, due to variations in fiber diameter and NMJ locations, is modelled at the structural level, created from Gaussian distributions simulating each variation. To model the 2^{nd} factor, a unique set of firing delays associated with the NMJ of each 'near' fiber in a MU is created using a Gaussian distribution for which the mean is zero and the standard deviation varies dependant of the degree of NMJ variability.

Finally, to form the MUP, the buffer related to 'distant' MFPs is added to 'near' MFP individual buffers and in this way the MUPs are modelled as composites of individual MFPs.

Motor Unit Recruitment and Firing Pattern Generation

An individual MU is not active until the level of muscle contraction is above some minimal level which is termed its recruitment threshold. Once a motor unit is active it remains active throughout a contraction as long as the force created remains above the recruitment threshold. Obviously as the level of contraction increases, more MUs become active. It is believed that larger motor units will be recruited, or become active, at relatively higher levels of muscle force compared to smaller motor units (Stashuk [1993]). This is known as the size principle and it is modelled by assigning MU recruitment thresholds based on MU territory sizes (Hamilton-Wright and Stashuk [2005]).

To maintain or increase the force generated by a muscle, MUs must fire repeatedly. Therefore during a sustained contraction each motor unit generates multiple MUPs, the collection of which is called a motor unit potential train. As the level of contraction increases, MUs firing rates, hence the number of MUPs per second in the trains increases. To simulate the firing behavior of a MU, its firing rate is calculated by directly relating it to the MU recruitment threshold and the defined level of contraction. Firing rates lie typically in the 8 pps - 42 pps range.

Inter discharge intervals (IDI) between sequences of MUPs in a train, the so-called MU firing patterns, are modelled as outputs from a Gaussian renewal point process. The mean is the reciprocal of the firing rate calculated for the specific MU based on the previously mentioned algorithm and the coefficient of variation is typically set to 0.15.

EMG Signal

All MUPs detected from active MUs along with their firing patterns are constructed to build a train, then each motor unit potential train is added to form the composite EMG signal. The resulting EMG signal represents the net voltage detected by the electrode positioned somewhere in the muscle.

3.1.4 Features used in this project

In order to relate electrophysiology to underlying anatomy, each should be characterized by appropriate features. Here, electrophysiology is measured by the simulated EMG signal and characterized by MUP features, whereas the underlying anatomy is represented by MU characteristics and described by the muscle model from which the EMG signal is generated.

MUP features	**Volume conductor factors**	**MU characteristics**
Area,	Distance to MU center,	Number of fibers,
Duration,	'Near' fiber count,	Averaged fiber density,
Thickness,	Mean distance to 'near' fibers,	Mean fiber diameter,
Size index,	Distance to closest MU fiber,	Stdv of fiber diameter.
Amplitude,	'Near' fiber count,	
MUP fiber count,	to mean distance ratio,	
Maximum	Needle Z position.	
acceleration,		
Maximum slope,		
Number of turns,		
Number of phases.		

Table 3.1: Specific electrophysiological features, volume conductor factors and anatomical characteristics assessed in this project. 'Stdv' stands for standard deviation.

The complete collection of features used in this project to describe different aspects of the MUs, MUPs and volume conductor factors (VCFs) are listed in Table.3.1.

In order to obtain the required features describing MUs, MUPs and VCFs, some modifications were applied to the existing simulator software. At the same time some new characteristics were introduced of which corresponding values were extracted from the simulator and the underlying muscle model. Extra routines were added to the simulator

software to define features describing distribution of fibers in each MU. In this regard, *mean fiber diameter* and *fiber diameter standard deviation* for each MU were calculated. As well, additional electrophysiological features were extracted; e.g. *size index* was calculated using Equation.2.1, having MUP *amplitude* and *thickness* from the existing software.

As mentioned before, volume conductor factors affect the EMG signal significantly, hence they have a great influence on the relationships between EMG features and muscle characteristics. Therefore, it is worthwhile to assess associations between these factors and MUP features at the same time as studying those between MUP and MU features. This helps us to understand the effect of VC factors on the inter-relationships between MUP and MU features. In this regard, additional algorithms have been designed to provide certain measurements characterizing volume conductor factors properly. The full collection of the VCFs used in this thesis, is mentioned in Table 3.1 in which the term 'distance' stands for the 'needle's tip distance' and 'Needle Z position' is the relative distance of the electrode to the endplate along the length of the fiber.

As a rule of thumb, it is believed that MUPs detected from MUs having fibers close to the electrode would be more informative compared to that of MUs with distant fibers. This means that as the needle moves towards the MU fibers, detected MUPs become more representative, hence relationships between MUP features and MU characteristics, if any, strengthen. Therefore, to assure maximized relationships between MUP and MU features, only significant MUPs should be assessed while non-representative MUPs should be excluded. The state of significance of a MUP is determined by the closeness of the MU to the electrode. In this regard, *near fiber count to mean distance (NFD) ratio* was introduced as an appropriate index VCF characterizing the closeness of a MU to the electrode and was measured by dividing *near fiber count* to *mean distance to near fibers* for each MU. It basically describes the group of fibers creating significant MFPs (MFP acceleration greater than 7.5 kv/s) and considers their average distance to the needle at the same time. Therefore *NFD ratio* reflects the closeness of the significant MU fibers to the electrode.

It should be further noted that in this thesis, three different kinds of fibers are defined as three separate features and they shouldn't be mistaken for one another. The MUP's *fiber count* is the number of peaks of the MUP's acceleration waveform happening beyond some threshold. MU's *number of fibers* is simply number of fibers located in the MU territory area.On the other hand, *near fiber count* is defined by counting fibers whose MFP's maximum acceleration is larger than 7.5 kv/s^2 (Stashuk [1999b]).

All of these features were measured for each performed EMG study characterizing different aspects of the muscle under study. They were then saved into files available for future referencing.

3.2 Relationship Discovery Technique

The anatomy model is combined with a particular needle configuration and a clinical EMG signal acquisition scheme to perform an EMG study and simulate an EMG signal. The acquired EMG signal along with the underlying muscle model, were then used to provide measurements of corresponding MUP features and MU characteristics. Based on the algorithm used to construct the model and generate the data, MU characteristics along with MUP features, related to a particular muscle state, are random variables (r.v.s). To find the associations existing among these r.v.s, (i.e., the goal of this project) different relationship discovery methods were applied. In this project correlation analysis and a pattern discovery technique were applied to the dataset to discover the inherent relationships between the features in the dataset. Through the following sections these two methods are introduced, their uses are justified, and their details are discussed.

3.2.1 Correlation Analysis

Correlation analysis is the most trivial and common method utilized when inherent patterns in a dataset are to be revealed. As it has been used by others (Zalewska and Hausmanowa-

Petrusewicz [1999, 2000], Cengiz et al. [2002], Pattichis and Elia [1999]), it was used in this project, to reveal possible relationships between an EMG signal and the underlying muscle.

Correlation is the degree to which two or more quantities are linearly associated. The extent of correlation between the values on the two axes, in a two dimensional plot, is quantified by the so-called correlation coefficient. The correlation coefficient is calculated easily and quickly and interpretations based on it are clear and straight forward. It possesses bipolar values providing helpful information regarding the direction of relatedness. If the correlation coefficient is positive, the two random variables (r.v.s) are directly related, i.e., as one increases the other increases. Whereas a negative correlation coefficient implies an inverse relationship (i.e., the value of one r.v. increases as that of the other r.v. decreases).

The correlation coefficient is an expected value that is used to describe the relationship between two continuous r.v.s $X(t)$ and $Y(t)$. It is also called as the *Pearson correlation coefficient* and is defined as r in the following format:

$$r = \frac{E(X,Y) - E(X)E(Y)}{\sqrt{\left(E(X^2) - E(X)^2\right)\left(E(Y^2) - E(Y)^2\right)}}, \tag{3.1}$$

where $E(X) = \int xP(x)dx$ is the expected value of random variable X and $E(X,Y) = \int_y \int_x xyP(x,y)dxdy$ is the joint expected value of random variables X and Y. It should be noted that Equation 3.1, calculates the correlation coefficient between two continuous-valued r.v.s. However, some of the features used in this project, such as fiber count and number of turns, are discrete-valued. To calculate the correlation when either one or both r.v.s are discrete-valued, the *Spearman Rank* correlation coefficient is calculated instead of *Pearson* coefficient.

To calculate the *Spearman coefficient*, each feature should be sorted first. Then the rank 1 is assigned to the first individual, 2 to the second, and so on, with the rank k assigned to the k^{th} individual. Note that tied ranks would get the average of the ranks.Having the two sets of ranked features, the *Spearman Rank* correlation coefficient, $\acute{r}$, is calculated from

the following formula:

$$\acute{r} = 1 - \frac{6 \sum D^2}{n^3 - n}, \tag{3.2}$$

where D is the difference between the *ranks* of the two features and n is the number of data points.

Throughout the thesis, *correlation coefficient* is termed as the coefficient of correlation measured between two features regardless of the nature of the data, (i.e., whether it is integer or real).

In this work the correlation coefficients between several MUP features and MU characteristics were calculated to examine the existence of two-sided linear relationships between them. Moreover, the calculated coefficients were evaluated by comparing with correlations of real data mentioned in the literature, for a full description see chapter 4.

3.2.2 Pattern Discovery Technique

Introduced in 1995 (Wong and Wang [1995]), this pattern discovery technique (PDT) is a conceptual clustering method based on an unsupervised learning strategy. It is applied to find the existing relationships or inherent patterns, in a dataset. A pattern is defined as a statistically significant association among two or more events in a problem domain, where an event is a feature holding a value. The significance is determined by the difference between observed occurrence in the dataset and the expected occurrence according to a default model. The co-occurrences of events are interpreted as non-random, hence describing a significant pattern, reflecting the nature of the dataset. In this project, the PDT was applied to datasets including several MU and MUP features, in order to find relationships between anatomical characteristics of muscle and features extracted from acquired simulated EMG signals. The PDT is applied to discrete-valued datasets, therefore the continuous-valued and large integer MU and MUP features used in this project, were first discretized.

Compared to correlation analysis discussed earlier, the PDT possesses three main advantages: 1) It finds associations between not only features but also their values, hence discovering discrete relationships existing in a dataset. By discrete it is meant that the relationships between features don't stand for the entire range of their values, rather the two features are associated when presenting some specific values. Therefore the PDT can reveal *any* relationship between features in dataset, including *discrete* or *whole range* . This capability is beneficial for the datasets used in this project, since the nature of its inherent patterns is quite unknown. Using PDT ensures that *any* pattern existing in these datasets would be revealed. 2) It discovers polythetic or high-order patterns[2] containing more than two features. The ability of discovering high-order patterns is important for the analysis of the complex datasets used in this thesis. Specifically this ability facilitates discovering relationships between two features regarding a third feature, such as an index feature (e.g. maximum slope). 3) This association detecting process is qualitative, since it determines whether a compound event is a significant association or not. In fact at the same time, it is quantitative because the significant level along with the probabilities are calculated and recorded.

In the upcoming sections, the terminology used in the technique and the methodology behind it is defined and described.

Notations and Terminology

Consider a dataset D containing M samples $D = \{S_1, S_2, ..., S_M\}$. Every sample in this dataset is described by N attributes, $S_j = \{X_1, X_2, ..., X_N\}$ each of which can be seen as a random variable. Each attribute, $X_i\,(1 \leq i \leq N)$, can hold a value from a set of discrete alphabets, $\alpha_i = \{\alpha_i^1, \alpha_i^2, ..., \alpha_i^{q_i}\}$ where q_i is the cardinality of the i^{th} attribute alphabet. Thus each of the M samples in the dataset D can be defined by $S_j = \{x_{1j}, x_{2j}, ..., x_{Nj}\}$

[2]High order patterns are those patterns regarding relationships among more than two factors. For example the propositions *Amplitude=high, Duration=low* and *Area=low* is a third order pattern.

which means that each attribute, $X_i\,(1 \leq i \leq N)$, has assumed a value from its alphabet and turned into $x_{ij}\,(1 \leq i \leq N and 1 \leq j \leq M)$. Assuming these notations, the following definitions have been defined:

Definition 1: A *primary event* of a random variable $X_i (1 \leq i \leq N)$, is an event in which X_i has assumed a value from its alphabet. Suppose X_i has taken its p^{th} value, then the p^{th} primary event associated with it would be $[X_i = \alpha_i^p]$ or in a simpler form x_{ip}. For the dataset D, there are v primary events such that:

$$\nu = \sum_{i=1}^{N} n_i$$

where n_i is the number of alphabets for each attribute. It is assumed that two primary events, x_{i1} and x_{i2} related to the same attribute X_i are mutually exclusive unless they are equal.

Definition 2: A set of primary events is called a *compound event*. Assuming s to be a subset of integers,$\{1, 2, ..., N\}$ then X^s is defined as a subset of sample S such that:

$$X^s = \{X_i | i \in s\}$$

Then x_j^s, the j^{th} realization of X^s is the j^{th} *compound event* associated with X^s. In this way x_j^s can be defined as the set of primary events created by the j^{th} realization of X^s:

$$x_j^s = \left\{X_i = \alpha_i^{pj} | i \in s, pj \in \{1, ..., n_i\}\right\}$$

A primary event is a *1-compound event* and a *k-compound event* contains k primary events of k different variables. Every sample in the dataset S_j $(1 \leq j \leq M)$ is an N-compound event.

Definition 3: A *sub-compound event* x_j^r is a subset of the compound event x_j^s such that, $\exists r, r \subset s$.

Definition 4: The *expected occurence* of a compound event x_j^s in D is its expected total under the assumption that the variables in X^s are mutually independent. The expected occurrence of x_j^s is denoted by $e_{x_j^s}$.

Definition 5: Assume T as a statistical significance test. If the occurrence of a compound event x_j^s is significantly different from its expected occurrence according to T, it is said that x_j^s is a *significant pattern* of order $|s|$.

Definition 6: The primary events related to a significant pattern x_j^s have *statistically significant association* or simply they are *associated*.

Detection of Patterns

For a compound event x_j^s, the probability under the assumption of independency between all of the variables in X^s, is called the expected probability. Considering a compound event, x_j^s, if the probability $P\left(x_j^s\right)$ is significantly different from the expected probability, then x_j^s is a significant pattern. This indicates that the primary events involved are likely to co-occur.

Let us denote the observed and expected occurrence of a compound event x_j^s as $o_{x_j^s}$ and $e_{x_j^s}$. The expected occurrence is calculated from the observed one using the following equation:

$$e_{x_j^s} = M \prod_{x_{ip} \in x_j^s} P\left(X_i = \alpha_i^p\right) = M \prod_{x_{ip} \in x_j^s} \frac{o_{x_j^s}}{M}, \tag{3.3}$$

where M is number of the samples in the training set D.

Instead of testing whether or not the probabilities are significantly different, the occurrences are compared with each other in a residual format. The residual concept related to an event is the difference between the occurrence of the event in the dataset D and its expected occurrence based on default probabilistic model. This residual concept is defined by:

$$d_{x_j^s} = o_{x_j^s} - e_{x_j^s}, \tag{3.4}$$

where $o_{x_j^s}$ and $e_{x_j^s}$ are the observed and expected occurrences of the compound event x_j^s in the dataset consequently.

Obviously, the absolute difference between the observed and the expected occurrence,

$|o_{x_j^s} - e_{x_j^s}|$, can't be used as a suitable distance measure, because it depends on the marginal probabilities. Practically the residual is standardized in the following format:

$$z_{x_j^s} = \frac{o_{x_j^s} - e_{x_j^s}}{\sqrt{e_{x_j^s}}}, \tag{3.5}$$

and further adjusted to better evaluate the relative size of the discrepancy between $e_{x_j^s}$ and $o_{x_j^s}$:

$$r_{x_j^s} = \frac{o_{x_j^s} - e_{x_j^s}}{\sqrt{c_{x_j^s}}} \tag{3.6}$$

where $c_{x_j^s}$ is the variance of $d_{x_j^s}$. The estimation of $c_{x_j^s}$ is presented in Appendix [A]. For details on the difference between these two residuals see Wong and Wang [1997], Wang [1997].

The adjusted residual $r_{x_j^s}$ and standard residual $z_{x_j^s}$ are distributed approximately as asymptotic normal distributions with zero mean and unit variance. Thus using either of them, if 95% is the desired level of confidence for the difference between the observed occurrence and the expected occurrence of an event, 1.96 should be used as the significance level (SL) by conventional criteria.

It is important to note that for patterns with small number of expected occurrences, noisy data points might pretend as significant events. Therefore to avoid this possible source of noise, a minimal number is assigned for expected number of occurrences as an essential condition. This means that, to investigate the significance of an event, its expected occurrence should be above this minimal threshold. On the other hand, the higher the order of the event is, the smaller is the expected number of that event. Given that, for small datasets, higher order patterns are not likely to be discovered.

For a compound event, x_j^s if the residual, $r_{x_j^s}$ exceeds 1.96 (95% SL) then the compound event is a significant pattern, meaning that the corresponding primary events are associated significantly and will likely occur together. In this case x_j^s is referred to as a positive significant pattern. Otherwise, if the true value of $r_{x_j^s}$ is less than -1.96, x_j^s is referred to as a negative significant event, which means that the corresponding primary events unlikely

co-occur. It is important to consider the negative significant patterns as well as the positive ones, since they contain information about the un-likeliness of the co-occurrence of the primary events and this is as important as the likeliness of co-occurrence.

To discover significant patterns inherent in a dataset, expected and observed occurrences are calculated first. Using those occurrences, the residuals are calculated by Eqn.3.4 and Eqn.3.6 one after another. Based on the selected significance level, the significant patterns for the dataset are then discovered. One can easily design an algorithm that exhaustively generates all the possible combination of the primary events at different orders and test their significance by calculating their adjusted residuals. In such an approach high-order contingency tables of different orders appear and make it computationally expensive. However, heuristics have been applied to the version of the PDT used in this thesis. Several algorithms have been used in which uninformative candidates for the significance test are eliminated at an early stage and prevented from further considerations. The details regarding the algorithms are not mentioned in this thesis and the reader is encouraged to read Wang [1997] for further information.

In this thesis, MU and MUP features were applied to the pattern discovery technique in order to detect the existing associations among them.

XOR problem, an example of PDT usage

As an illustration, an XOR problem will be considered in this subsection. Assume an XOR database including 1000 samples. Each primary event of this database occur 500 times, i.e., the marginal probability for each attribute is 0.5. Hence the number of expected occurrences of the 3-compound event $[A = T, B = T, C = F]$ is $0.5 \times 0.5 \times 0.5 \times 1000 = 125$. To test whether this compound event is significant, the standardized residual is then calculated. Suppose that the observed occurrence of this 3^{rd}-order event in the XOR database is 300. The standardized residual is calculated from Eqn.3.5 to be 15.65. This value is larger than 1.96, 95% SL, therefore this compound event is a 3^{rd}-order significant

pattern in the database. Meaning that $[A = T, B = T, C = F]$ is significantly different from its expectation or that the three primary events of $[A = T]$, $[B = T]$ and $[C = F]$ are associated, hence likely co-occur.

Patterns of Different Order

A real world database contains patterns of different order. The existence of higher order patterns does not guarantee the existence of lower order patterns and vise versa. Generally if x_j^k is the k^{th} order pattern in dataset D, any t^{th} order $(t < k)$ sub-compound event of x_j^k may not necessarily be a pattern in dataset D, i.e., occurring throughout dataset D. Thus the associations between events can only be found by testing the candidates of that order. For example in the XOR problem mentioned above, $[A = T, B = T, C = F]$ is a 3^{rd} order pattern but none of the sub-compound events $[A = T, B = T]$, $[A = T, C = F]$ and $[B = T, C = F]$ may be necessarily significant patterns. Thus synthesizing higher order patterns from the lower order ones is not possible. Also, even if all of $(k-1)^{th}$ order sub-compound events of x_j^k are significant patterns, the x_j^k itself may not be a significant pattern. Therefore, whether or not a compound pattern is significant, can not be determined through testing its sub-compound patterns.

Discretizing continuous data

Most of the features used in this project are continuous-valued random variables, yet PDT is applied to discrete-valued datasets. Thus to find the relationships between these features, they were discretized before being applied to the PDT relationship discovery methods.

There are different procedures for discretizing continuous-valued datasets. The MU and MUP features studied in this project, are random variables. Therefore the discretization scheme used in this project ensures that the transformed data is also a random variable; providing same probabilistic properties as prior to discretizing. This is done through maximizing the entropy of a discretized feature (i.e., equalizing the probability of each bin

happening). The simplest way to do this, is to bucket an equal number of data points in each bin, hence the probability of all bins becomes equal, however if two values are the same, it is not that easy to perform the strategy. To overcome this, an optimized algorithm explained by Chau [2001] was used to properly discretize the continuous features.

Chapter 4

Experiments and Results

In order to study the relationships between MUP features and anatomical characteristics of MUs, the analysis methods described in Chapter 3, were applied to data created using the model described earlier. In this chapter, relationships between MUP and MU features are revealed and discussed. Moreover specific experimental situations along with relevant assumptions are outlined.

4.1 Preliminary Experimental Routines

To obtain the required data, i.e., proper MUP and MU features, the simulation model described in 3.1 was used. A concentric needle configuration was constructed as part of the model. The parameters of the simulator were set such that a normal muscle was modeled.

Physicians tend to acquire multi-sets of EMG signals from each studied muscle using a minimal number of insertions, to maximize patient comfort. To model this, two different insertion sites were simulated for each muscle modeled. Knowing that as the needle moves away from endplate region, MFPs become more disperse in time than those detected near the neuromuscular junction, one site was selected near to the endplate region while the

other was chosen a bit further away, to include all varieties of MUP shapes. For each site, the needle location along the length of the fiber was determined relative to the NMJ region and was selected randomly from a range. Three different recordings were performed during each insertion. Actually the electrode was repositioned twice after each insertion. In Fig.4.1 this is shown by arrows from $a1$ to $b1$ and from $b1$ to $c1$. The second set of recordings regarding the other insertion site are marked as $a2$, $b2$ and $c2$ in Fig.4.1. Therefore for each muscle, six EMG signals were simulated.

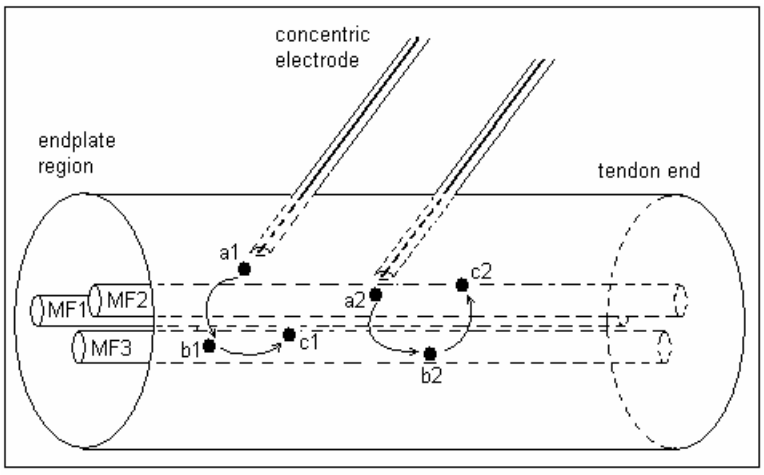


Figure 4.1: Needle was inserted twice in each muscle. For each insertion site, three EMG recordings were obtained, adding up to 6 EMG signals for each muscle model.

To ensure the same MUs were not resampled during the three EMG signal acquisitions at each insertion site, the minimal distance between adjacent needle locations was calculated. To do this, extra sets of experiments were performed, in which common numbers of MUs involved in EMG signals acquired from different needle locations were calculated. Figure 4.2 shows that 4 mm is the proper minimal distance between adjacent detection sites to avoid data redundancy due to sampling the same MU during a series of EMG studies.

For consistency across various sets of experiments, superficial recording sites close to

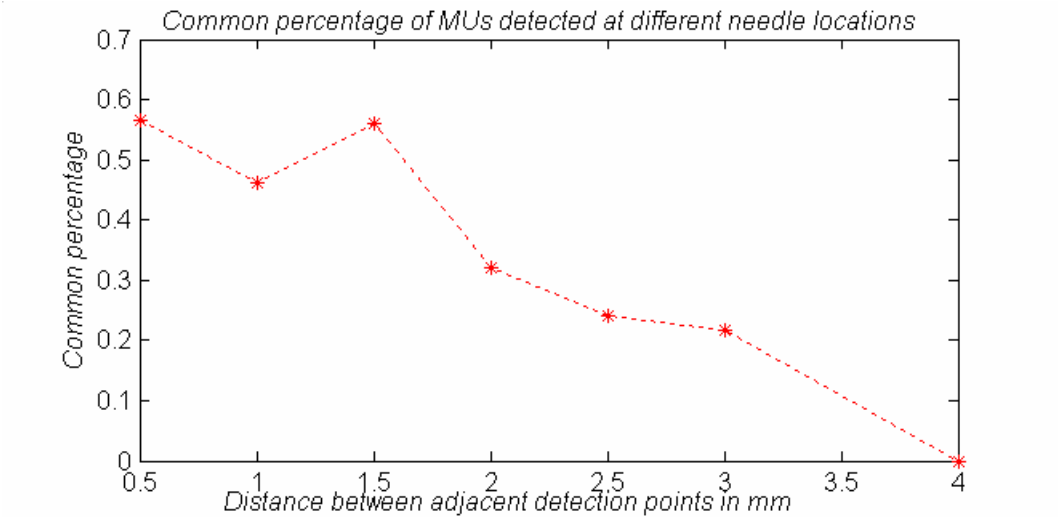


Figure 4.2: Common percentage of the MUs spotted during EMG acquisition in adjacent needle locations.

the muscle surface were avoided (Falck et al. [1995]). Therefore the triangle was always centered at the center of the muscle cross section area. After selecting the positions of the electrodes, the simulator was executed, creating 6 EMG signals for each muscle modelled. For the research of this thesis, 15 muscle models were generated in total. These EMG signals were then resolved into their constituent motor unit potential trains using a clinically accepted decomposition technique, provided through the DQEMG software. This decomposition routine includes several signal processing and pattern recognition steps. The details regarding this technique are beyond the scope of this thesis and the reader is advised to see Stashuk [1999a, 2001] for more information. The decomposed MUPs were then used to extract specific electrophysiological features. Similarly the required anatomical characteristics were obtained for the associated MUs from which the MUPs were detected.

On average 13 MUPs were detected from each EMG signal and having 15 muscle models with 6 EMG signals in each, results in over 1000 MUPs in total. The extracted MU and

MUP features were applied to the analytical routines described in the previous chapter. However, in each case, particular statistics were defined to ease final referencing. The results for each section are included as well.

4.2 Correlation Analysis

Correlation analysis was performed as the most common and simplest relationship discovery method. It was conducted between features in each group as the inter-grouped association. The results are presented in the following tables.

MUP Features	dur	thick	SI	amp	FC	acc	slp	turn	phase
area	0.44*	-0.43^{-}	0.88*	0.96*	0.52	0.70	0.84	0.72	0.59^{+}
duration (dur)		0.16^{-}	0.60*	0.33*	0.07	0.19	0.25	0.26	0.33*
thickness (thick)			-0.18^{-}	-0.58*	-0.63	-0.58	-0.63	-0.58	-0.45^{-}
size index (SI)				0.78*	0.42	0.52	0.65	0.65	0.68^{+}
amplitude (amp)					0.6	0.79	0.92	0.77	0.57^{+}
fiber count (FC)						0.78	0.7	0.68	0.49
max acceleration (acc)							0.93	0.82	0.67
max slope (slp)								0.81	0.77
number of turns (turn)									0.75

Table 4.1: Correlation coefficients between MUP features used in this work.

In Table 4.1, the coefficients marked with an asterisk are the ones consistent with correlation coefficients mentioned in the literature (Cengiz et al. [2002]), between MUP features obtained from real EMG signals. However, non-consistent[1] correlations are either marked

[1]An explanation is presented in section 4.2.3.

MU Features	Average density	Mean fiber diam	Stdv of fiber diameter
Number of fibers	0.19	0.73	0.16
Average density		-0.14	-0.01
Mean fiber diameter			0.1

Table 4.2: Correlation coefficients between MU characteristics used in this work.

VC Factors	DClose	NFibC	MDNFib	StdvDNFib	NFibDis	Zposition
Distance to MU center,	0.81	-0.77	0.25	-0.12	-0.76	-0.07
Distance to closest fiber (DClose),		-0.66	0.71	-0.75	-0.68	-0.03
Near fiber count (NFibC),			-0.17	0.21	0.98	0.06
Mean distance to near fibers (MDNFib),				-0.42	-0.26	0.03
Stdv of distances to near fibers (StdvDNFib),					0.49	-0.02
Near fiber count to mean distance ratio (NFibDis).						0.05

Table 4.3: Correlation coefficients between volume conductor factors used in this work.

with a hyphen or a plus sign determining whether this value is much smaller or larger than the real one, correspondingly. The unmarked coefficients associate to correlation coefficients not found in the literature.

There is not a high correlation between most of the MU features mentioned in Table

4.2, except for *mean fiber diameter* and *number of fibers* in the MU. This is well explained, since *mean fiber diameter* in each MU is determined based on the size of that MU in the muscle model used. In fact the high correlation coefficient is consistent with specific assumptions considered in the simulation model.

The correlation coefficients presented in Table 4.3 agree with specific definitions for individual volume conductor factors. Note that the needle position along the length of the fiber *Z Postion*, is not related to any other VCF. Unfortunately no correlation analysis between volume conductor factors was not found in the literature.

4.2.1 Volume Conductor Factors as Indexes

As mentioned in Chapter 2 and 3, VCFs alter the associations existing between MUP features and MU characteristics. Therefore to make sure the relationships between MUP and MU features are maximized, detected MUPs should be filtered based on specific VCF values. This aspect of volume conductor factors were further investigated by studying the triple associations existing among three groups of MU features, MUP statistics and VC factors.

A special procedure was determined to use correlation analysis in discovering associations between two features relative to a third selection or index feature. In this so-called Sorted-Correlation scheme, the correlation coefficient between two features was calculated relative to a third feature, i.e., index feature. As the very first step in this procedure data points were sorted based on the values of index feature. The correlation coefficient between the two features was calculated and stored. Data points associated with the smallest value of the index feature were then eliminated from the dataset. The correlation coefficient between the remaining set of data points was then calculated and stored. Again data points corresponding to the next smallest value of the index feature were eliminated and then the correlation coefficient was calculated and stored. This procedure was repeated till no data points were left. However, values of the correlation coefficient during the last couple

steps were non-stable and noisy, hence those steps regarding the last 10% of the data were not stored. The correlation coefficients stored during each step, were then plotted in one figure presenting the trajectory of correlation between two features relative to a third index feature.

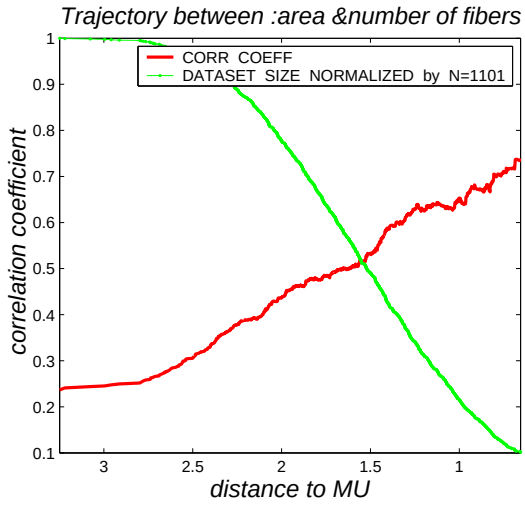
Since MU size is well addressed by MUP *area*, further analysis was conducted to investigate the effect of volume conductor factors on the association between *area* and number of fibers in the MU. Therefore several correlation trajectories were performed. Figures 4.3 and 4.4 present the alterations in correlation coefficients between *area* and *number of fibers* in the MU relative to the values of several VCFs. Note the increase in the correlation coefficient as MUPs detected from distant MUs (Figures 4.3(a) and 4.3(b)) and those MUs with fewer near fibers (Figure 4.3(c)) get eliminated from the dataset. As well those MUPs detected from MUs without any near fiber, were excluded from the analysis performed for Figures 4.3(c) and 4.3(d). That's why the number of data points for these figures is different from the others.

On the other hand, as shown in Figure 4.4 neither *needle Z position* nor *stdv of near fiber distance* affect the inter-relationships between *number of fibers* in the MU and *area*.

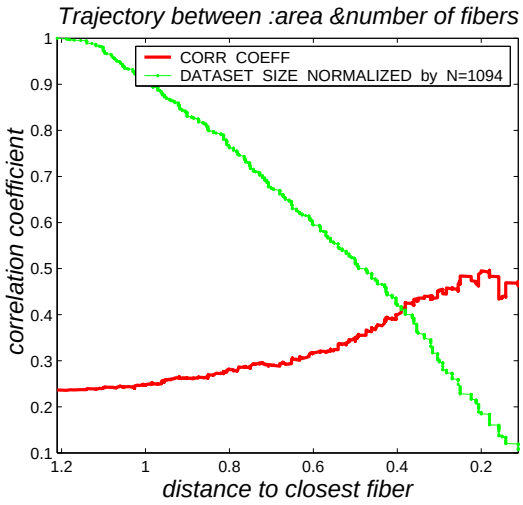
From Figure 4.3, it is obvious that the correlation trajectory between *area* and MU size starts from a higher point and elevates the most, when *near fiber count to mean distance (NFD) ratio* is used as the index feature. This index VCF was defined as the division of *near fiber count* and *mean distance to near fibers* in section 3.1.4. It actually emphasizes the presence of near fibers when they are close to the needle, and suppresses their presence when they are far away. Since the correlation trajectory shown in Figure 4.3(d) behaved in the most desired way, we studied the index VCF, *NFD ratio* a bit further.

4.2.2 Equivalent MUP Index

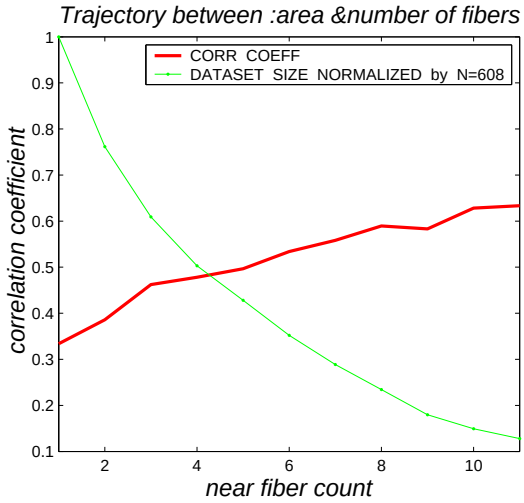
The Sorted-Correlation procedure was performed between MU size and some MUP features: area, duration, thickness and amplitude. The VCF *NFD ratio* was used as the index



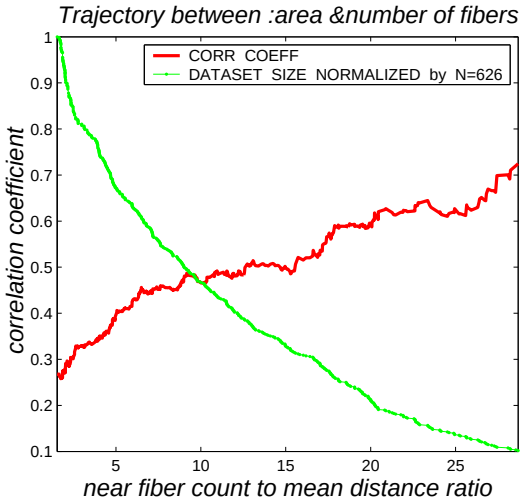
(a) index = *Needle distance to MU center*



(b) index = *Needle distance to closest fiber*



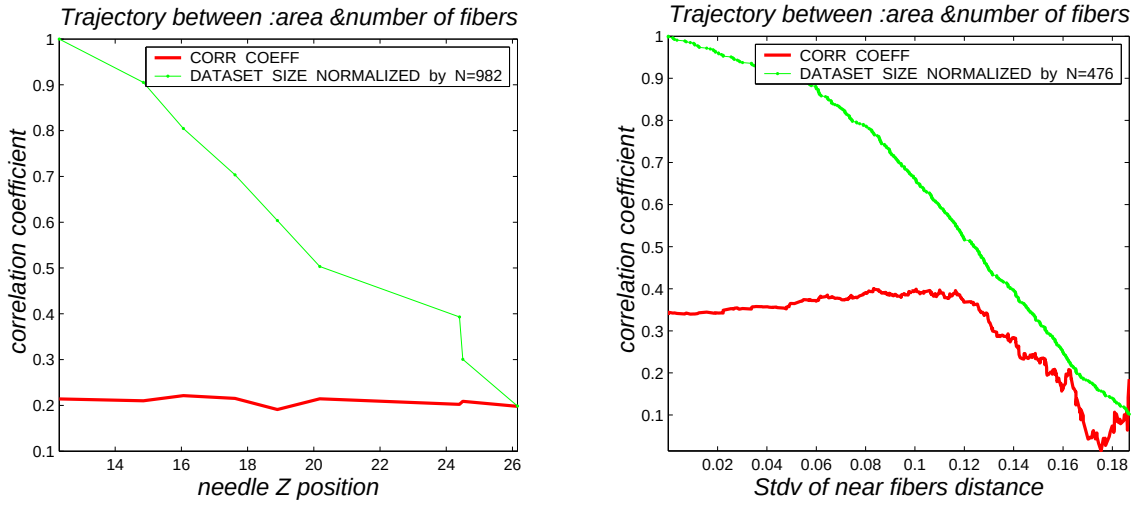
(c) index = *Near fiber count*



(d) index = *Near fiber count to mean distance ratio*

Figure 4.3: Correlation trajectories between *area* and MU size versus different volume conductor factors.

feature. Since *NFD ratio* reflects the closeness of important MU fibers to the electrode, it was used as the index in selecting significant MUPs. Corresponding trajectories are shown



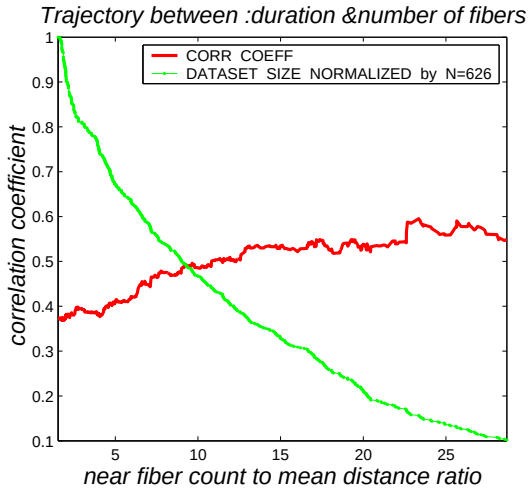
(a) (b)

Figure 4.4: Correlation trajectories between *area* and MU size versus different volume conductor factors.

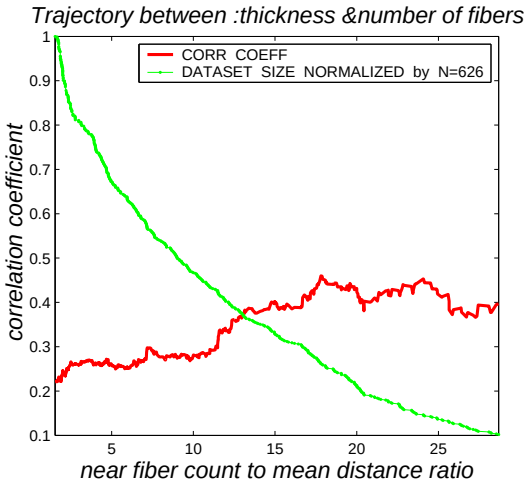
in Figure 4.5.

Based on Figure 4.5 the correlation coefficient between *number of fibers* in the MU and MUP features do increase as this volume conductor factor increases. This means that MUPs from MUs with high *NFD ratio* reflect the size of the MU better than other MUPs.

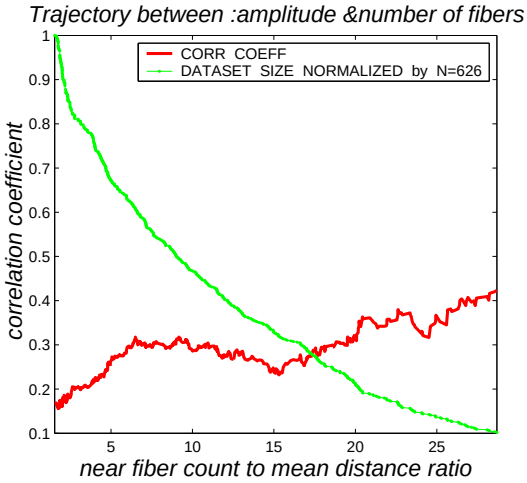
Since it is desired to maximize the relationships between anatomical characteristics of the muscle and features of a detected EMG signal, *NFD ratio* threshold above which correlation coefficients between MU and MUP features are maximum possible, should be found. However, in real clinical settings information regarding volume conductor factors is not available to the physician since the position of the needle relative to active muscle fibers in MU is not known. Given that the only source of information in real clinical settings is the detected EMG signal, an equivalent MUP feature was found so that the index feature could be measured and studied from real muscles. This MUP feature was then used as the index feature instead of VCFs in investigating the inter-relationships between MUP features and MU characteristics.



(a) Trajectory between *duration* and MU Size



(b) Trajectory between *thickness* and MU Size



(c) Trajectory between *amplitude* and MU Size

Figure 4.5: Correlation trajectories between several MUP features and MU size versus *near fiber count to mean distance ratio.*

MUP index features	maximum slope	maximum acceleration	Size index
r	0.74	0.64	0.62

Table 4.4: Correlation coefficients between MUP index features and VCF *NFD ratio.*

Several correlation trajectories were performed between possible MUP indexes and *NFD ratio*, to find the appropriate MUP feature that is highly correlated with this VCF. As recorded in Table 4.4 the correlation coefficient between *maximum slope* and *NFD ratio* was the highest among the three candidates. It should be noted that the correlation trajectories were performed for MUs with different sizes, hence MU's *number of fibers* was used as the index feature in those trajectories. Figure 4.6 demonstrates that the correlation trajectory between *maximum slope* and *NFD ratio*, is independent of MU size; the correlation coefficient along the trajectory line stays constant for different MU sizes. This means that *maximum slope* is the best index feature and could be used to select significant MUPs that are most[2] representative of the MUs from which they are recorded.

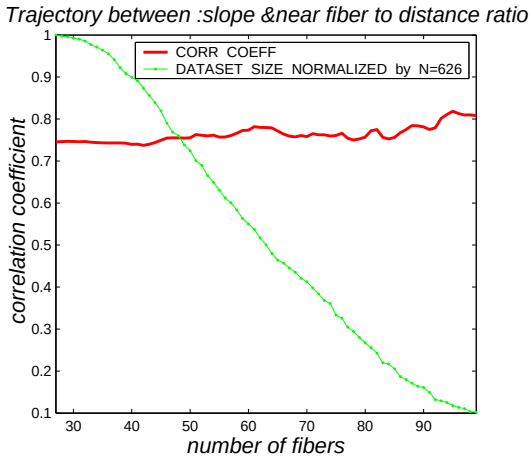


Figure 4.6: Correlation trajectory between *NFD ratio* and MUP *maximum slope* relative to *number of fibers (NumFib)* in the MU.

Likewise, others (Barkhaus and Nandedkar [1996], Stashuk [1999b]) have used *maximum slope* along with *maximum acceleration* as a criteria or index parameters for selection of sharp MUPs, i.e., those MUPs being used for assessment when presented in clinical EMG recordings. They believed that existing relationships between MUP features and MU

[2]being informative about the size of the MU

characteristics grow stronger as the MUP gets sharper and more significant. Stalberg *et al.* (Stalberg et al. [1995]) have used 0.3 v/s for *maximum slope* as the inclusion threshold for significant MUPs. This implies that MUPs with *maximum slopes* greater than 0.3 v/s are more representative of their MUs, hence stronger relationships should exist between features of those MUPs and anatomical characteristics of associated MUs. We tried to study this number as part of this work.

4.2.3 MUP Inter-Grouped Trajectories

Before studying the cross-grouped relationships existing between two groups of MUP features and MU characteristics, the inter-grouped relationships of non-consistent coefficients were further studied. Sorted-Correlation procedure was conducted between MUP features mentioned in Table 4.5. Corresponding trajectories are mentioned in the following figures.

Based on trajectories shown in Figures 4.7 and 4.8 the correlation coefficients between MUP features will be close to what has been mentioned in the literature, if the data is filtered properly. This means that if MUPs with specific slopes are excluded from the dataset, e.g. lower than 0.8 v/s for Figure 4.7(b) or lower than 0.6 v/s for Figure 4.8(a), correlation coefficients would be the same as the ones in the literature.

In the following sections we tried to investigate slope's proper threshold above which cross-grouped and inter-grouped relationships maximize.

4.2.4 Correlations Between MU and MUP Features

Based on the previous sections we were confident enough to use *maximum slope* throughout the rest of the thesis, as the index feature in investigating the associations between MU characteristics and MUP features. Correlation trajectories were performed between MU features and MUP features mentioned in Table 4.6, where *maximum slope* was used as the selection criteria for significant MUPs. From now on the term **slope** is used instead of **maximum slope** for convenience.

Coupled MUP features	$\mathbf{r_l}$	$\mathbf{r_t}$
thickness and area	0.47	-0.43
thickness and duration	0.47	0.16
thickness and SI	0.60	-0.18
thickness and phase	-0.15	-0.45
phase and area	-0.04	0.59
phase and SI	-0.07	0.68
phase and amplitude	0.05	0.57

Table 4.5: Comparison between correlation coefficients mentioned in Table 4.1 and corresponding values in Cengiz et al. 2002. $\mathbf{r_l}$ is the correlation coefficients mentioned in the literature and $\mathbf{r_t}$ is non-consistent correlation coefficients mentioned in this thesis.

MU feature	**MUP feature**
number of fibers	area,duration,thickness,size index,amplitude
average fiber density	area,duration,thickness,size index,amplitude
mean fiber diameter	area,duration,size index,amplitude,number of turns, number of phases
Stdv fiber diameter	duration,size index,number of turns,number of phases

Table 4.6: Features applied to Sorted-Correlation procedure.

Using simulation models, Nandedkar et al. (Nandedkar et al. [1988c]) reported a correlation coefficient of 0.24 between *number of fibers* in the MU and MUP *area*. It is interesting that exactly the same association ($r = 0.24$) exists for our data, Figure 4.9(a). However, the trajectory plot demonstrates an increase in the association between MUP *area* and MU size when the MUPs get more representative, i.e., *maximum slope* increases. Note that after MUPs with *slope* values lower than 0.6 v/s are excluded, the correlation

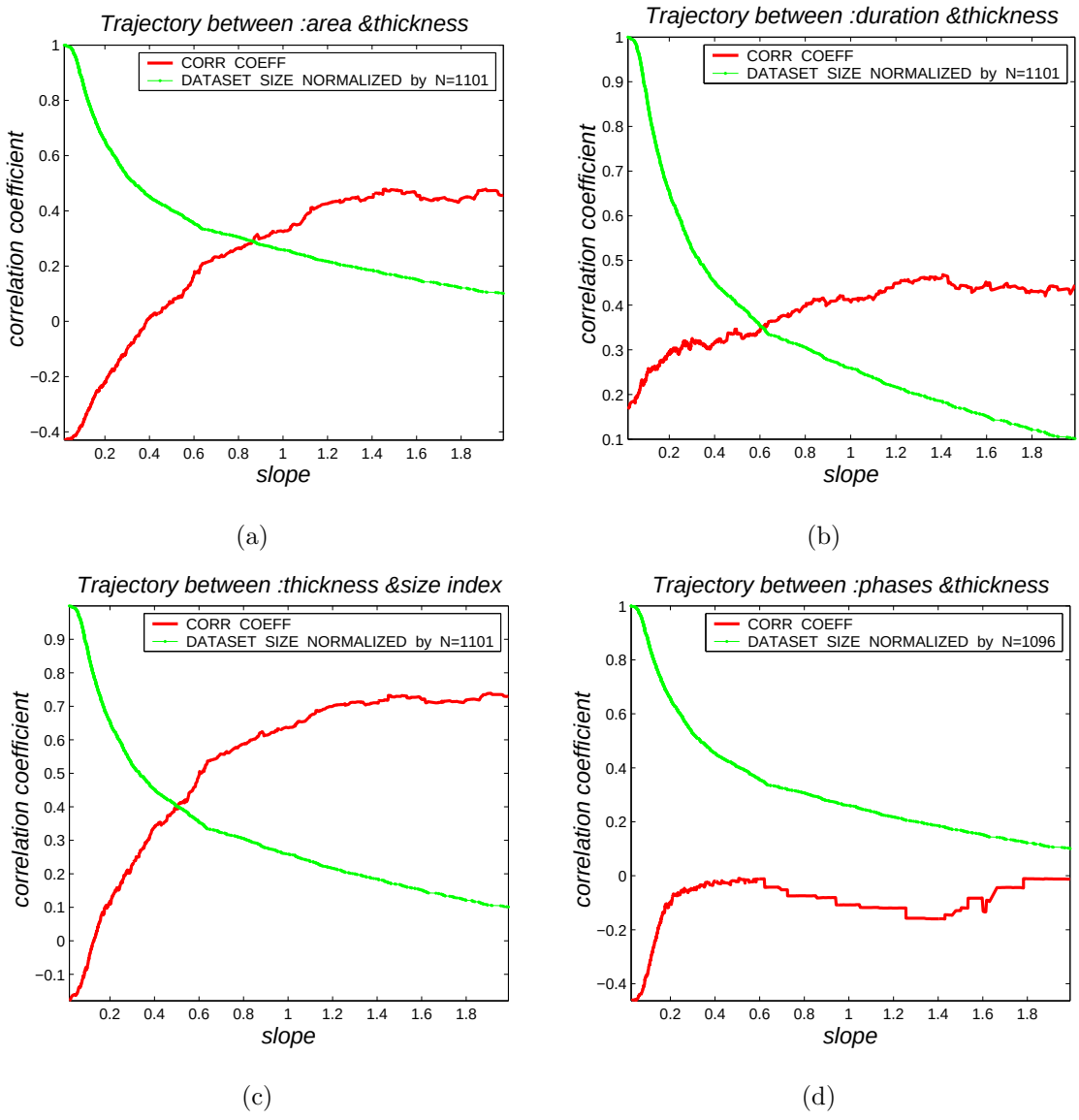


Figure 4.7: Correlation trajectories between MUP features mentioned in Table 4.5 versus *maximum slope*.

coefficient stays constant independent of *slope*.

On the other hand, MUP *duration* has been mentioned by others (Stalberg and Karlsson [2001a]) as representative of MU size. However, simulation studies have suggested that *duration* measurements from only simple MUPs (i.e., MUPs with numbers of phases less

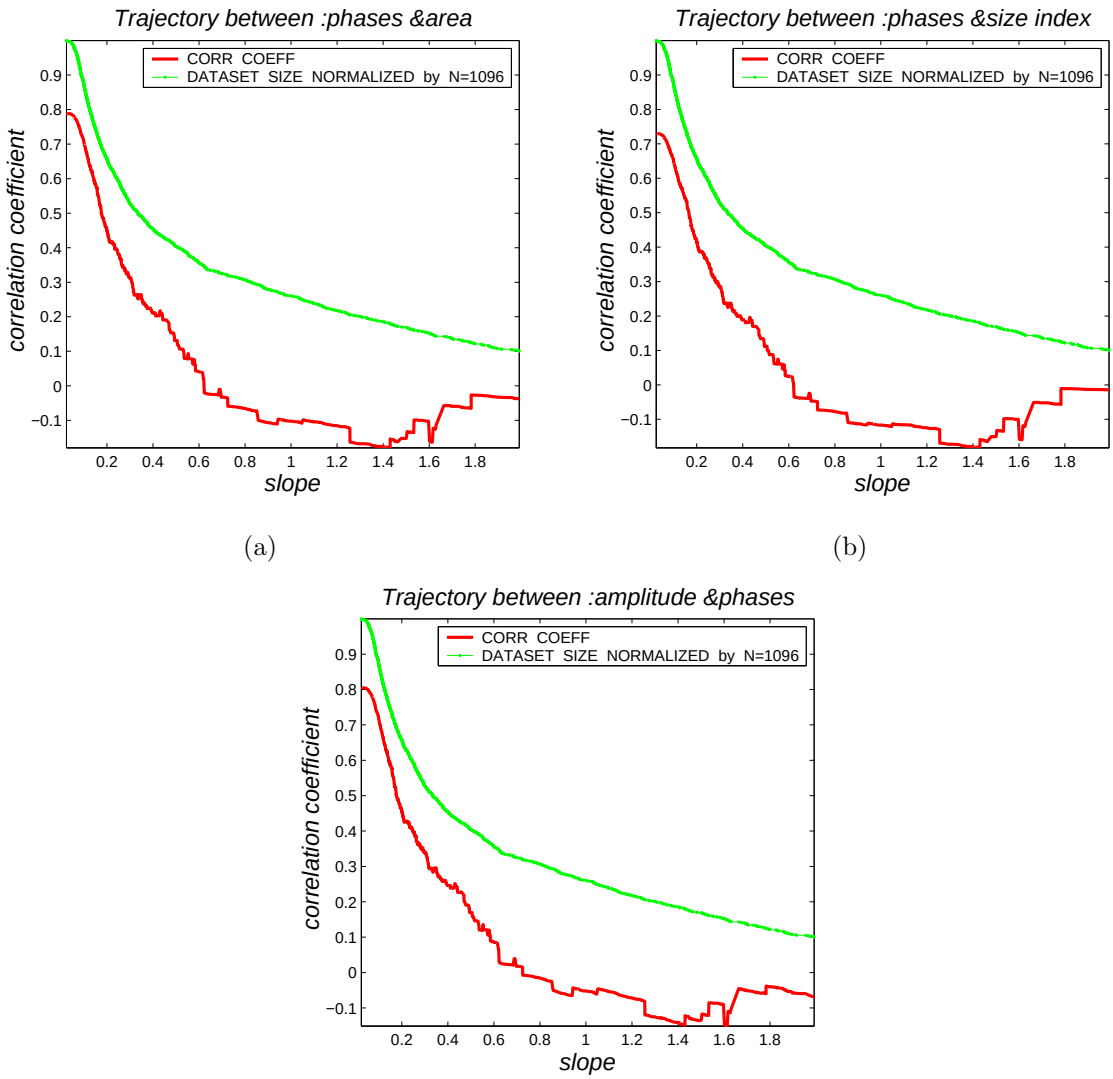
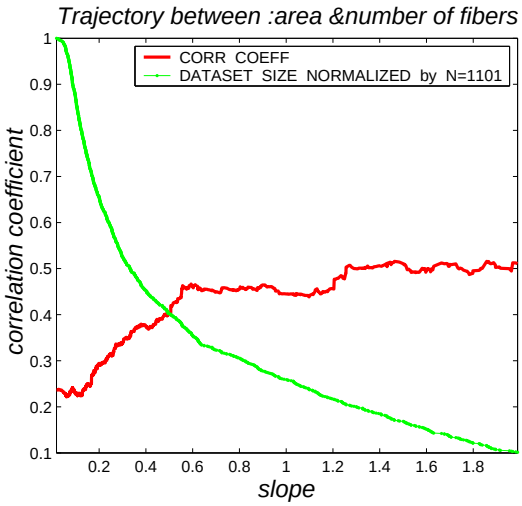
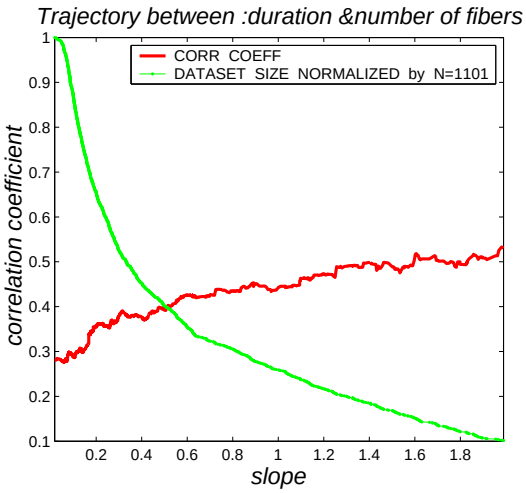


Figure 4.8: Correlation trajectories between MUP features mentioned in Table 4.5 versus *maximum slope*.

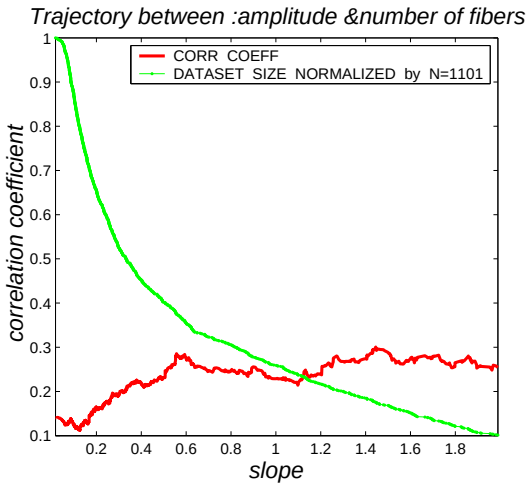
than 5), contain information about the size of the MU (Nandedkar and Sanders [1989]). In EMG signals acquired from normal muscles most of the MUPs are simple shaped, therefore it is expected that MUP *duration* correlates with MU size or number of fibers in the MU. Since a simulated model for a normal muscle was used in this work the kind of correlation



(a) Trajectory between *area* and MU Size



(b) Trajectory between *duration* and MU Size



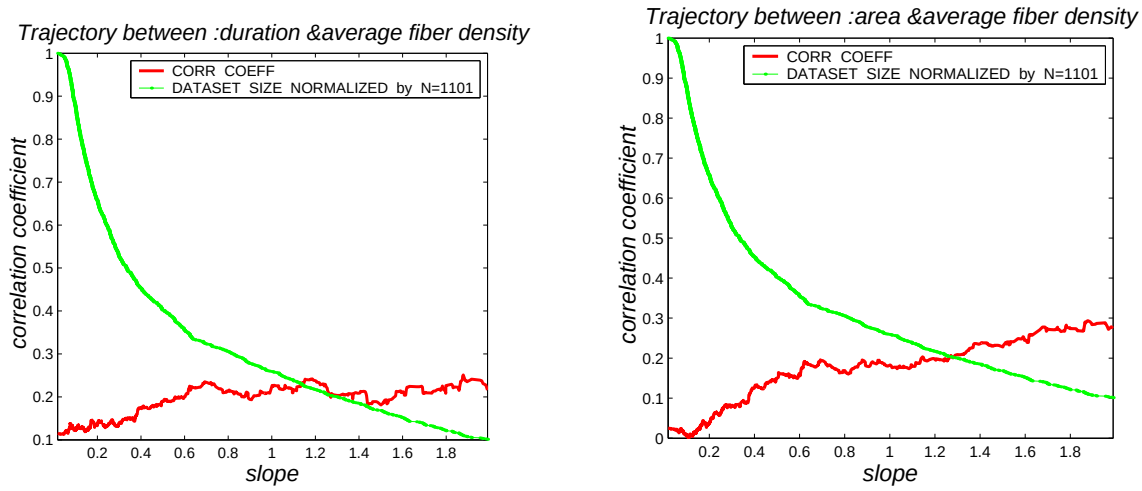
(c) Trajectory between *amplitude* and MU Size

Figure 4.9: Correlation trajectories between several MUP features and MU size versus *maximum slope*.

shown in Figure 4.9(b), is totally reasonable and hence reassuring. Additionally the starting point correlation coefficient ($r = 0.31$) as shown in Figure 4.9(b) is the exact correlation coefficient mentioned between *duration* and *number of fibers* in the MU by Nandedkar et al. [1988c]. Note that this value also increases if MUPs with higher *maximum slope*

are considered. Part (c) of the same figure, displays the lower correlation between MUP *amplitude* and MU size as expected (Nandedkar et al. [1988c]) which again improves with proper MUP selection.

Similar correlation analysis were conducted for *size index* and *thickness*. The correlation trajectories (not shown here) between these two features and MU size were found to be similar to those presented for *area* and *duration* correspondingly.



(a) duration

(b) area

Figure 4.10: Correlation trajectory between *average fiber density* of MU and some MUP features relative to *maximum slope*.

Correlation trajectories were performed between MU *average fiber density* and MUP features mentioned in Table 4.6. All correlation trajectories had a knee point at *slope* = 0.6 − 0.7 v/s except for *thickness* whose correlation was below 0.2 and independent of *slope*. Corresponding correlation trajectories for *duration* and *area* are shown in Figure 4.10.

As well correlation trajectories were conducted between MU's *mean fiber diameter* and some MUP features, Figure 4.11. *Mean fiber diameter* was determined to be reflected by *size index* (Figure 4.11(a)) and *area* (Figure 4.11(b)) for MUPs with *slopes* greater

than 0.6 v/s, ($r = 0.3$). *Duration* was also a good determinant of *mean fiber diameter*, however the MUPs have to be filtered further, Figure 4.11(c). Specifically MUPs with slopes greater than 0.85 v/s had durations correlated ($r = 0.32$) with distributions of the fiber diameters in the MUs. *Amplitude*, on the other hand, was not related to *mean fiber diameter* ($r = 0.14$) but yet the correlation coefficient increased a bit as MUPs with slopes lower than 0.6 v/s were excluded. Interestingly neither *number of turns* ($r = 0.02$) nor *number of phases* ($r = -0.08$) were correlated with *mean fiber diameter* and the de-correlation was constant along the trajectory line.

MUP *duration* is not related ($r = 0.06$) to stdv of fiber diameter as shown in Figure 4.12(a). Since a normal muscle was simulated, mostly simple MUPs were detected. Thus the effect of MU size on MUP *duration* was more pronounced than that of temporal dispersion caused by variation in fiber diameters. However, for high *slope* MUPs, the correlation between MUP *duration* and stdv of fiber diameter increases slightly. This may be because the closeness of the fibers contributing to sharp MFPs to the MUPs, have exaggerated the effect of fiber diameter variation on *duration*.

Number of turns and *number of phases* are MUP parameters representing variations in fiber size, hence their inter-relationship with *fiber diameter standard deviation* were investigated by applying the Sorted-Correlation procedure. The results are shown in Figure 4.12. Correlation coefficient between *number of turns* and stdv of fiber diameter increases as the MUP's *maximum slope* increases, Figure 4.12(b). In contrary with MUP turns, *number of phases* is not correlated with *stdv of fiber diameter*, Figure 4.12(c). The correlation coefficient along the trajectory reaches up to 0.25 when *number of turns* is used, while for *number of phases* the maximum value for correlation coefficient was 0.11, even for highly significant MUPs, (i.e., sharp MUPs). This explains why *number of phases* is not used for discrimination between myopathic and normal muscles. Variation in fiber diameter increases as an early sign in myopathies, therefore *number of phases*, which is not sensitive to this factor, can not characterize the diseased tissue sufficiently to be used as a

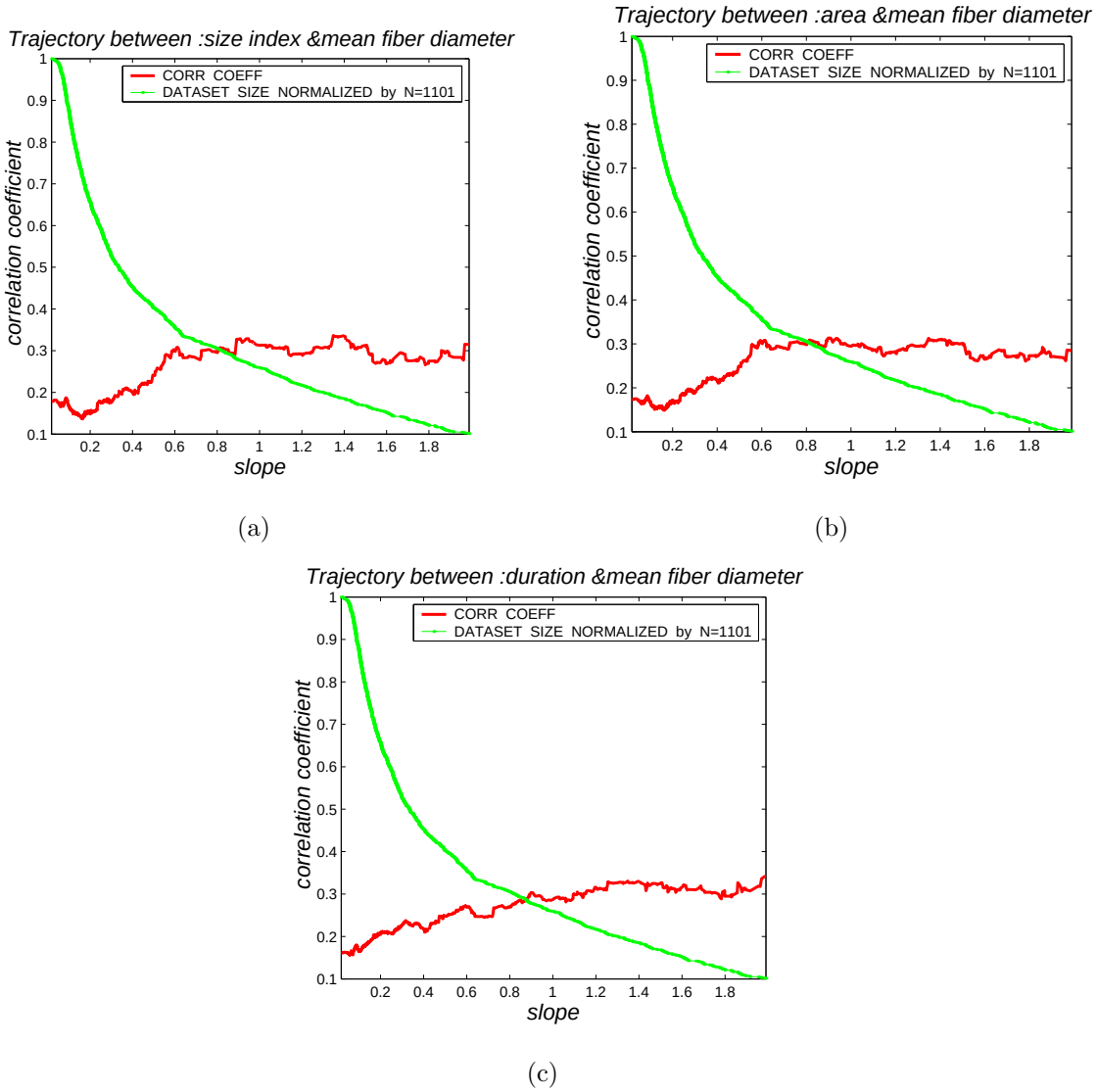
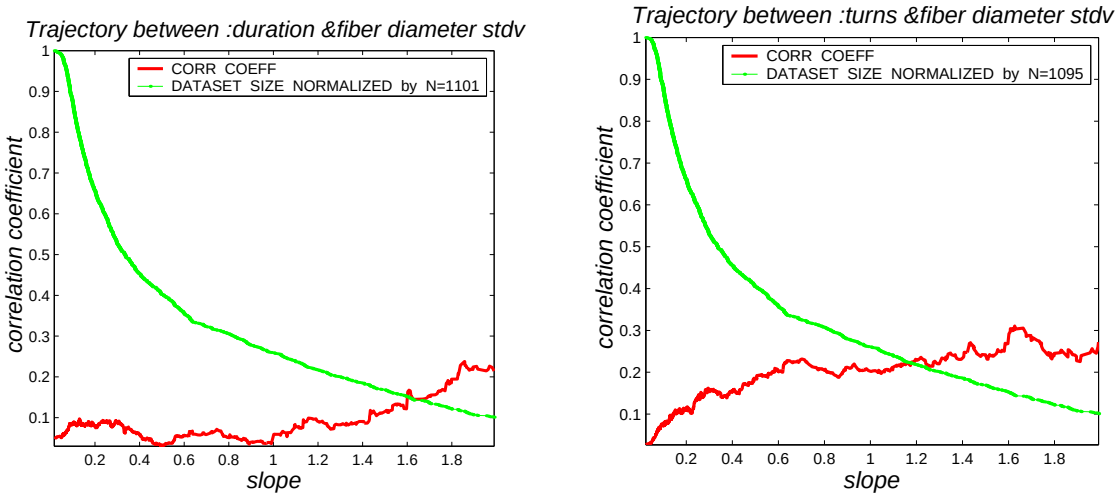


Figure 4.11:

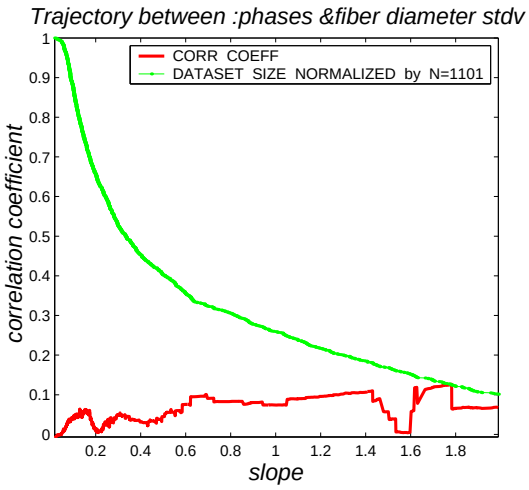
discriminant. Note the knee point at $slope = 0.7$ v/s in Figure 4.12(b).

So far several correlation trajectories between MUP features, MU characteristics and volume conductor factors have been presented. Most of the trajectories, included a *slope* threshold or knee point above which correlation coefficients were roughly constant and stabilized. For convenience the stabilized correlation coefficients between MUP features and



(a) duration

(b) number of turns



(c) number of phases

Figure 4.12: Correlation trajectories between several MUP features and MU's *standard deviation of fiber diameter* versus *maximum slope*.

MU characteristics are summarized in Table 4.7. Slope thresholds at which the correlations are reached is also provided.

Based on the slope thresholds mentioned in Table 4.7, the appropriate threshold for *slope* is suggested to be 0.6 v/s. Therefore, to obtain maximum possible for relationships

MU feature	**MUP feature**	$\mathbf{r_s}$	$\mathbf{slope_{thr}}$
number	area	0.48	0.6
of	duration	0.45	0.7
fibers	thickness	0.45	0.7
	size index	0.48	0.6
	amplitude	0.3	0.6
average	area	0.25	0.65
fiber	duration	0.25	0.6
density	thickness	0.1	-
	size index	0.2	0.63
	amplitude	0.17	0.65
mean	area	0.3	0.6
fiber	duration	0.32	0.85
diameter	size index	0.3	0.6
	amplitude	0.14	0.6
	number of turns	0.02	0.2-1.5
	number of phases	-0.08	0.2-1.4
Stdv	duration	0.06	-
fiber	size index	0	-
diameter	number of turns	0.25	0.7
	number of phases	0.07	-

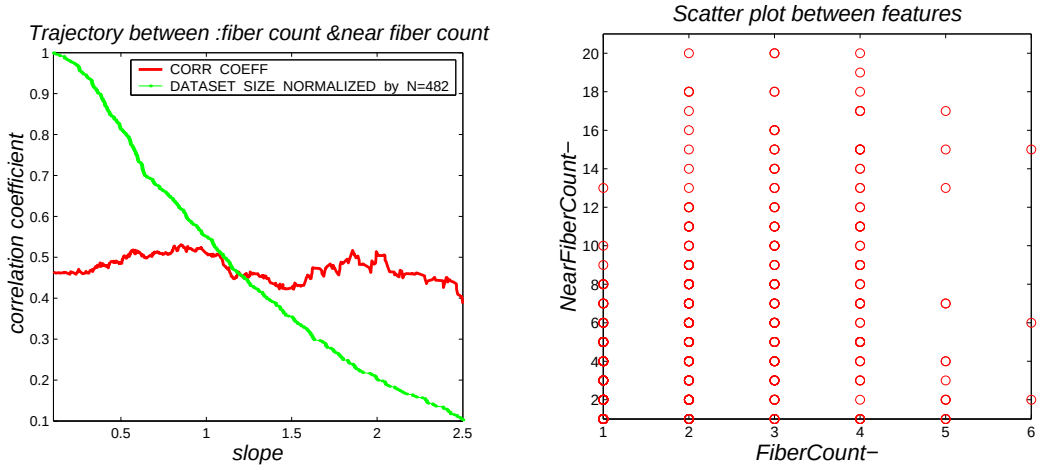
Table 4.7: Final results of correlation analysis between MU characteristics and MUP features. $\mathbf{r_s}$ stands for stabilized correlation coefficient and $\mathbf{slope_{thr}}$ is the slope at which $\mathbf{r_s}$ is reached.

between MUP features and MU characteristics, only MUPs with slopes above 0.6 v/s should be considered.

Correlations Between *MUP Fiber Count* and *Near Fiber Count*

In addition to finding interrelationships between MUP features and MU characteristics, some interesting relationships were revealed between *MUP fiber count* and *near fiber count*. Since *near fiber count* is related to MU characteristic, the relationship is really an interrelationship between a MUP feature and MU characteristic. Moreover since the relationship was interesting it is worthy enough to be reported and mentioned as part of the thesis.

Near fibers as defined earlier, are those fibers whose MFP's maximum acceleration is above 7.5 kv/s. This volume conductor factor represents well the fibers located in 500 μm proximity of the needle electrode, since *near fiber count* were highly correlated ($r = 0.94$) with *local fiber count*[3] and the relationship (trajectory not shown here) was independent of the slope. Having said that, Sorted-Correlation scheme was performed between VCF *near fiber count* and *MUP fiber count*.



(a) Correlation trajectory

(b) Scatter plot

Figure 4.13: Correlation analysis performed between *MUP fiber count* and VCF *near fiber count*.

Figure 4.13(a) demonstrates that the correlation coefficient between *near fiber count* and *MUP fiber count* stays at $r = 0.45$ independent of the slope. To properly interpret

[3]counts of those fibers located in the 500 μm proximity of the needle

the correlation trajectory, corresponding scatter plot for *near fiber count* versus *MUP fiber count* were plotted, Figure 4.13(b).

4.3 Pattern Discovery Results

From the studies proposed up to now, it is evident that not only are there inter-relationships among MU characteristics, MUP features and volume conductor factors, but also there exist cross-relationships among these features. Thus to find existing relationships in the dataset a multi dimensional problem should be solved.

The pattern discovery technique (PDT) was applied to the features in order to detect the possible high order significant associations among the features and their values. Yet only third order patterns were studied in this thesis and investigating higher order patterns was left for future studies.

Before using PDT, the continuous features were discretized into 5 bins, since PDT is applied to discrete-valued datasets. 5 was chosen since each bin can then stand for a linguistic equivalent translation as shown in Table 4.8. The discretization routine used for discretzing the features ensured maximum Entropy for the quantized data.

Bin Number	1	2	3	4	5
Linguistic Equivalent	low	below average	average	above average	high

Table 4.8: Continuous features are discretized prior to apply PDT.

To avoid the possible noise introduced in case the expected number of events are small and to make sure the significant residuals are accurate, a minimal number of 5 was assumed for the expected occurrences.

4.3.1 How to Interpret the Discovered Patterns

PDT was applied to discretized features. Since third order patterns were intended to be studied, three features were selected one of which was *maximum slope* representing significant MUPs. In order to visualize the discovered significant patterns in a useful manner, residual values related to each discovered patterns were plotted in a colored form. The positive residuals were marked as green while red was chosen as the proper label for negative residuals. Darkness of the color showed the magnitude of the residual, i.e., the bigger the residual in magnitude, the darker the color was. Figure 4.14 presents the significant patterns discovered between *area* and MU size (number of fibers in the MU) when MUPs with certain *slope* values were excluded. This filtering strategy is the same as the one used for the correlation analysis.

By comparing the four plots in Figure 4.14 with each other, the alteration in alignment of the green cells on the diagonal can be noticed. At the same time, the red cells move to the corners as low *slope* MUPs get cleared from the dataset. This behavior confirms the increase in correlation coefficient between *area* and MU size as low *slope* MUPs get excluded (Figure 4.9(a)).

Obviously the information content of significant patterns should be presented in a more quantitative manner rather than that mentioned in the previous paragraph. To do this the following statistics were introduced:

Algebraic sum of significant adjusted residuals (ALS) which is calculated by adding adjusted residuals of significant patterns algebraically considering the signs of the residuals, i.e., adding up the magnitude of positive residuals while subtracting the magnitude of negative residuals.

Absolute sum of significant adjusted residuals (ABS) which is computed by adding up the magnitude of adjusted residuals regardless of the signs of the residuals.

Diagonal correlative accumulated significant adjusted residuals (DCA) in which

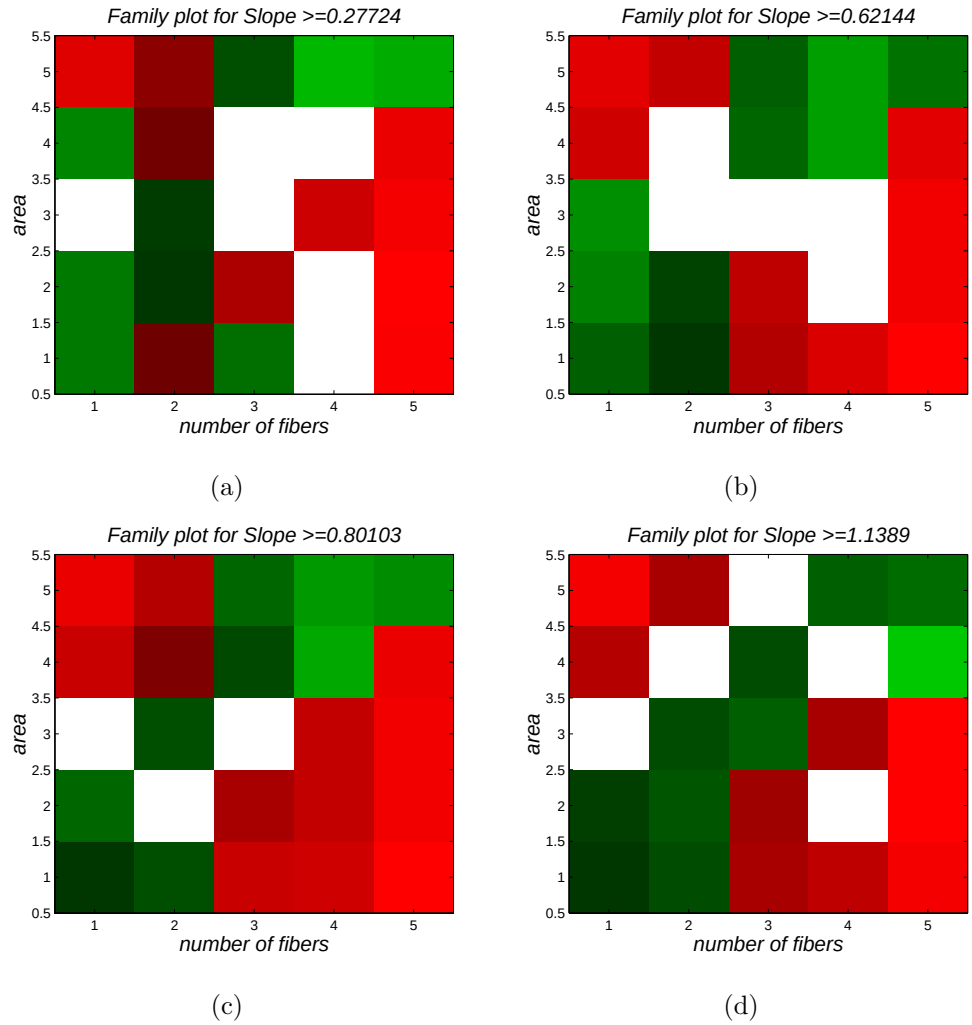


Figure 4.14: Significant patterns existing between *area* and *number of fibers* in the MU, for filtered MUPs based their slope.

residual values of the cells located along the $x = y$ line are added considering their signs unchanged, whereas the signs of residual values of the cells in corners are reversed Figure 4.15(a).

Quadratic correlative accumulated significant adjusted residuals (QCA) in which residual values of the cells located in the second and forth quadrants are added con-

sidering their signs unchanged, whereas the signs of residual values of the cells in first and third quadratics are inverted, Figure 4.15(b).

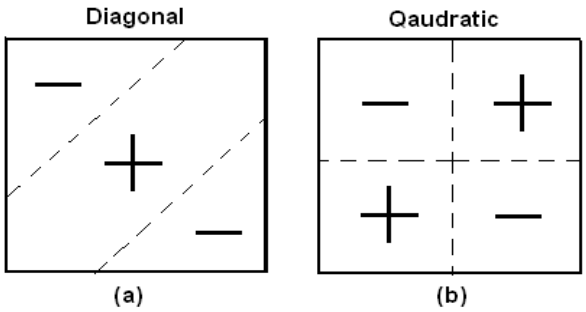


Figure 4.15: The sign of the residual values are considered relative to their location in QCA and DCA computation.

Note that in order to calculate QCA, the number of discretization bins is required to be odd. Given that 5 has equivalent linguistic form for interpretation reasons, 10 is selected as the suitable number of discretization units.

A similar scheme as the Sorted-Correlation procedure was used, through steps of which, the proposed statistics were calculated for the filtered data. As the very first step PDT was applied to the whole dataset and the four statistics were calculated for the discovered patterns. For the next step, low *slope* MUPs were eliminated and the PDT was applied for the new dataset. The four statistics were computed again. This scheme is called Sorted-PDT and it gets repeated till 10% of the MUPs remains. The statistics are then normalized by the maximum possible number of patterns and plotted to create corresponding trajectories. Figure 4.16 demonstrates these four statistics for patterns discovered between *area* and *number of fibers* in the MU when *maximum slope* is used as the index.

Based on Figure 4.16 it seems that ABS is the upper bound for the alignment factor, i.e., ABS is reached if all green cells get aligned on the diagonal and all red cells get concentrated in the corners, Figure 4.17. ALS, on the other hand, doesn't consider the

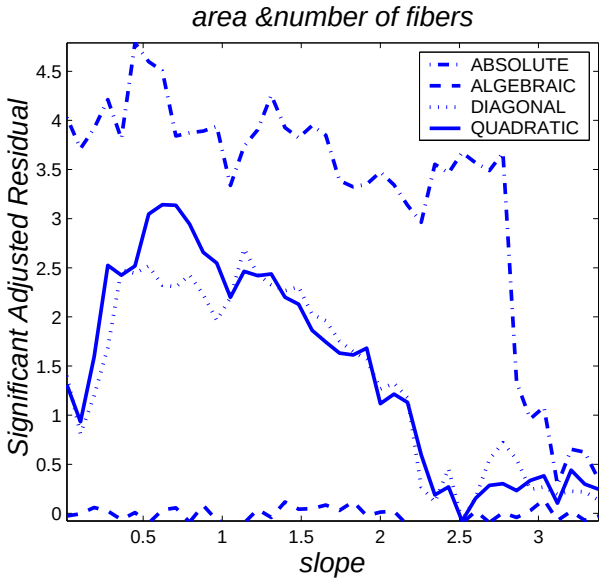


Figure 4.16: PDT trajectories of patterns discovered between *area*, *number of fibers* in the MU and *slope*.

alignment or concentration of the cells at all. Therefore it is not a statistic. However, the other two statistics, DCA and QCA reflect the concentration or alignment of the green cells (significant patterns with positive residuals) on the diagonal and concentration of red cells (significant patterns with negative residuals) off the diagonal in the corners very well. Therefore DCA and QCA were selected as the proper statistics to quantize the information content of significant patterns discovered between MUP features and MU characteristics.

4.3.2 Discovered Patterns Among MUP Features and MU Characteristics

To discover inherent patterns among MUP features and MU characteristics, the Sorted-PDT algorithm was applied to MUP features and MU features mentioned in Table 4.9. MUP *slope* was also included as the third index parameter. When significant inherent patterns were discovered, third order patterns were analyzed further and proper statistics

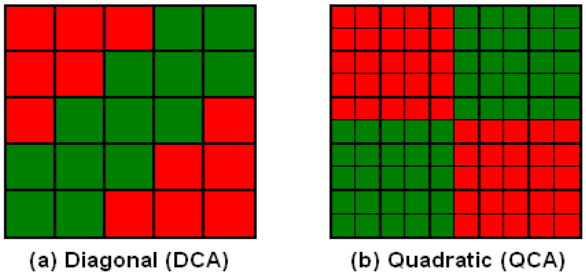


Figure 4.17: Specific configuration of the cells for which ABS is reached. DCA in (a) stands for diagonal correlative accumulated adjusted residuals, while QCA in (b) is quadratic correlative accumulated adjusted residuals.

(DCA and QCA) were computed through the Sorted-PDT algorithm. Computed values for DCA and QCA were plotted as trajectories which looked similar to correlation trajectories presented earlier. The knee points for these statistics and the *slope* values at which the statistics reached the point, for patterns between features listed in Table 4.9 are mentioned in Table 4.10.

MU feature	**MUP feature**
number of fibers	area,duration,thickness,size index,amplitude
average fiber density	area,duration,thickness,size index
mean fiber diameter	area,duration,size index,amplitude,number of turns number of phases
Stdv fiber diameter	duration,size index,number of turns,number of phases

Table 4.9: Features applied to Sorted-Correlation procedure.

The Sorted-PDT algorithm was applied to *number of fibers* in the MU, *slope* and some MUP features. Corresponding trajectories are illustrated in Figure 4.18. You can find the knee point in all of the trajectories, which means selective MUPs are good representatives of their MUs. QCA in Figure 4.18(a) suggests that by filtering the MUPs based on $slope = 0.6$

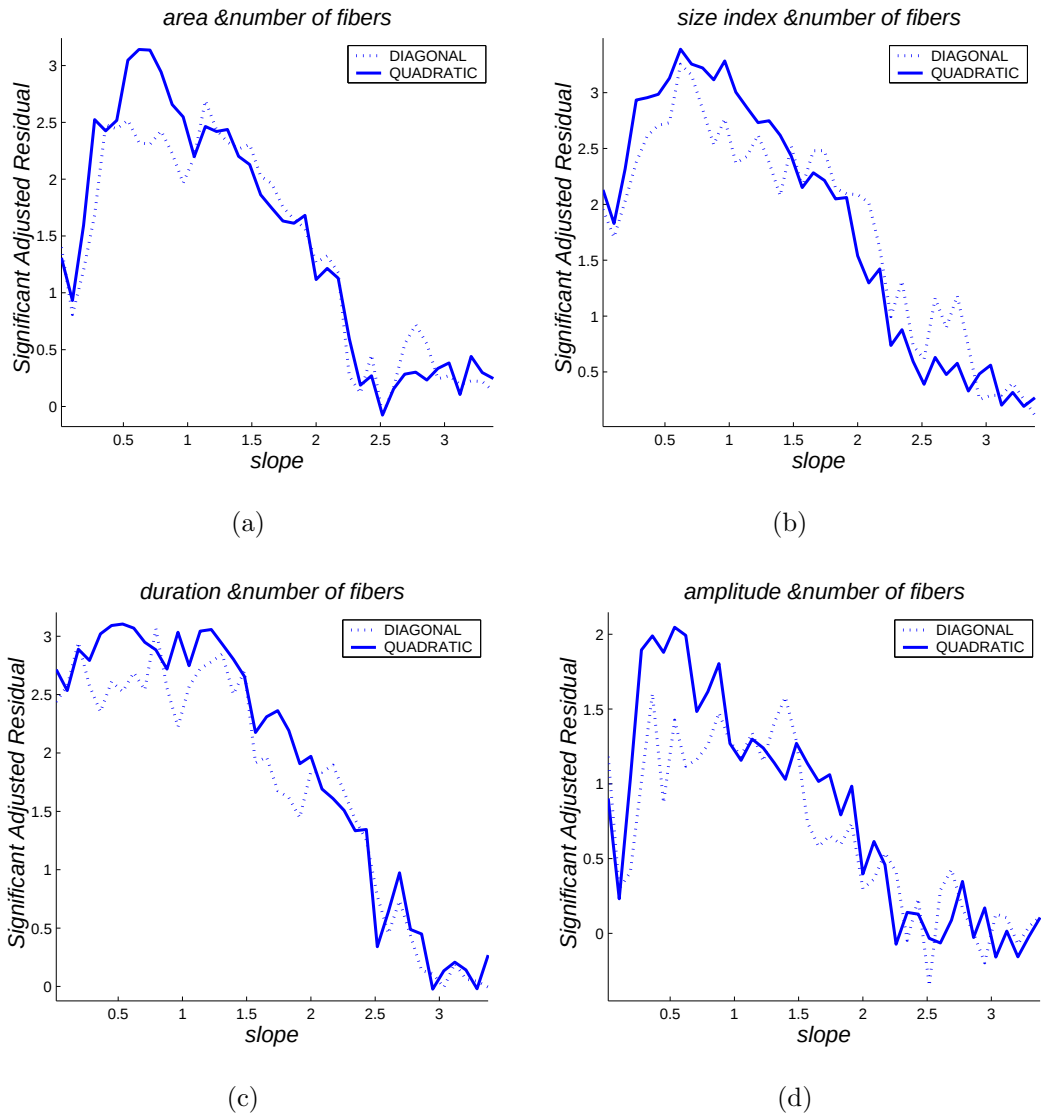
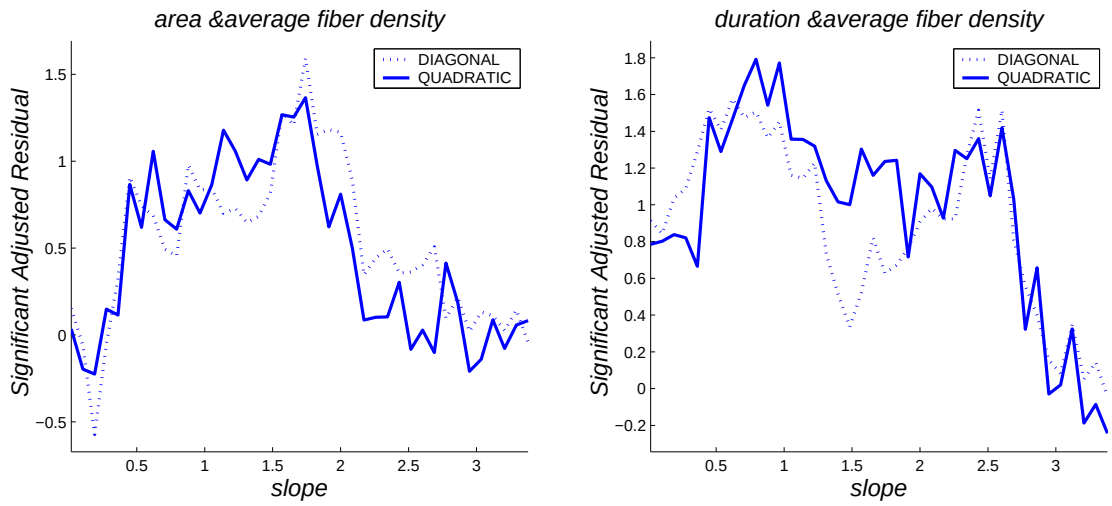


Figure 4.18: DCA and QCA trajectories of significant patterns existing among *slope*, *number of fibers* in the MU and some MUP features.



(a) (b)

Figure 4.19: PDT trajectories for discovered patterns between MUP features and MU average fiber density.

v/s residual values for discovered patterns between *area* and *number of fibers* in the MU would be more significant. On the other hand, DCA suggests a lower threshold for *slope* (0.4 v/s) and the residual values stay constant for a longer duration after the knee point, compared to QCA. In Figure 4.18(b) QCA doesn't contain the peak present in the previous plot and DCA drops earlier compared to the Figure 4.18(a). There is an insignificant increase for DCA statistic right in the beginning of Figure 4.18(c) suggesting that, including all MUPs, even dull MUPs with low slopes, *duration* is a good representative of MU size. However, QCA suggests that to maximize the relationship between *duration* and MU size, MUPs with slopes lower than 0.37 v/s should be excluded. On the other hand, *amplitude* is not a good feature characterizing MU size, since DCA and QCA are the lowest compared to other trajectories, Figure 4.18(d). Even so, there is a knee point at slope of 0.42 v/s, meaning that the weak association between *amplitude* and *number of fibers* in the MU increases as MUPs are filtered based on that *slope*. Trajectory between *thickness* and MU size was similar to *size index*, except the statistics were lower.

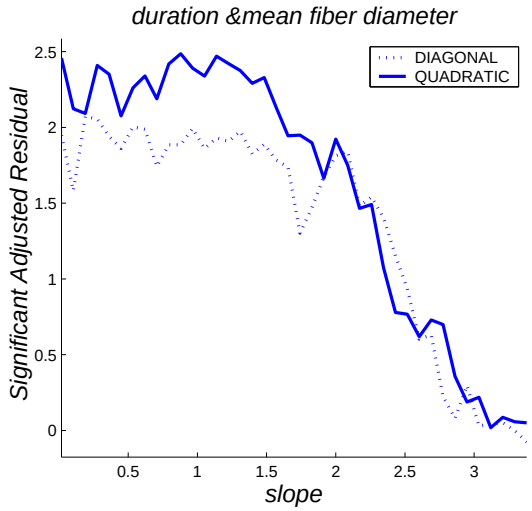


Figure 4.20: PDT trajectories for discovered patterns between *duration* and *mean fiber diameter* in the MU.

Significant patterns were discovered between *average fiber density* and MUP features. As shown in Figure 4.19, *duration* is a better determinant for *average fiber density*, since residuals are more significant. The knee point is better presented by QCA statistics in both plots, rather than DCA. The associations with *average fiber density* for *size index* and *thickness* were also analyzed (trajectories not shown here). The trajectory for *size index* was similar to what is shown for *area*. The trajectory for *thickness*, suggested that if all MUPs are considered, *thickness* is a good representative of *average fiber density*.

Sorted-PDT algorithm was applied to *mean fiber diameter* and some MUP features. DCA and QCA were the highest for patterns discovered between *duration* and *mean fiber diameter*, shown in Figure 4.20. It is interesting to see that DCA and QCA are constant independent of *slope*. Figure 4.20 suggests that when all MUPs are included, *duration* is a good determinant for *mean fiber diameter*, i.e., no filtering is needed. Patterns between MUP *area* and *mean fiber diameter* were discovered as well as the ones between *size index* and *mean fiber diameter*. For *area* and *size index*, DCA statistics suggest that MUPs with

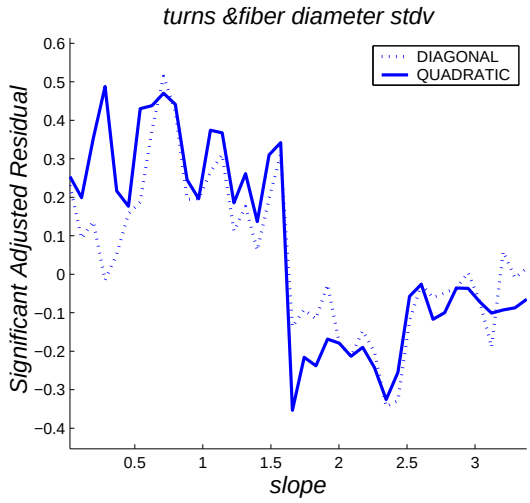


Figure 4.21: PDT trajectories for discovered patterns between *number of turns* and stdv of fiber diameter in the MU.

slopes lower than 0.5 v/s and 0.46 v/s should be excluded, whereas QCA suggest slope thresholds to be equal to 0.7 v/s and 0.6 v/s.

The only MUP feature providing information about the *stdv of fiber diameters* in the MU is *number of turns* of the detected MUP. Nevertheless the average adjusted residual values is not high and it drops drastically after MUPs with slopes lower than 1.5 v/s have been excluded, Figure 4.21.

The knee points for DCA and QCA statistics have been summarized and mentioned in Table 4.10. The slope threshold at which the knee points happen are also included.

MU feature	**MUP feature**	**QCA_{thr}**	**$slope_{qca}$**	**DCA_{thr}**	**$slope_{dca}$**
number	area	3.0	0.6	2.5	0.4
of	duration	2.82	0.25	2.82	0.25
fibers	thickness	2.19	0.7	2.58	0.7
	size index	2.9	0.3	2.7	0.46
	amplitude	-	-	1.6	0.4
average	area	0.87	0.48	0.87	0.48
fiber	duration	1.44	0.47	1.55	0.6
density	size index	0.80	0.49	-	-
mean	area	1.8	0.70	1.67	0.5
fiber	size index	1.9	0.6	2.2	0.46
diameter	amplitude	1.09	0.35	1.3	0.35
	number of turns	-	-	0.4	0.6
	number of phases	0.35	0.4	0.26	0.5
Stdv fiber	duration	0.1	0.4	-	-
diameter	number of turns	-	-	0.29	0.6

Table 4.10: Final results of PDT obtained from discovered patterns between MU characteristics and MUP features. **$slope_{qca}$** stands for the slope at knee point for QCA. **$slope_{dca}$** is the slope at at knee point for DCA. The ones hyphenated didn't contain a knee point.

Chapter 5

Discussion

The main goal of this work was to reveal associations existing between MU characteristics and MUP features. To reach this goal, a Sorted-Correlation algorithm was introduced through which correlation analyses were performed between MUP features and MU characteristics. Additionally a pattern discovery technique was utilized to discover the inherent patterns in the dataset. Results were presented thoroughly in Chapter 4. Corresponding discussions are provided in this chapter.

Resembling Real Data

The simulated data used in this thesis resembled real data acquired in a clinical settings. The inter-grouped correlation analysis, Table 4.1, matched the ones reported in the literature for features obtained from real muscles. In real clinical environments, EMG is acquired during low level contractions, and window triggering is used to extract single MUPs from the whole signal. The physician then measures the shape of the MUPs through eye-balling the signal.

In contrast, the features used in this project were extracted from a decomposed simulated EMG signal. Based on the similarities between our data and real data, the following two facts were confirmed: 1) DQEMG techniques work properly in extracting the infor-

mation content of EMG signals. 2) The simulation model works suitably in creating data resembling real data.

Correlation analysis

Knowing that volume conductor factors affect the relationships between MUP features and MU characteristics, we tried to find an index VCF for which the inter-relationships between features were maximized. For this purpose, *near fiber count to mean distance ratio* was selected as the best index feature, for which the correlation trajectory between *area* and MU size was maximized among all other VCFs (see Figures 4.3 and 4.4).

Since information regarding volume conductor factors are not available in real clinical studies, an equivalent MUP index was found. To find the appropriate MUP feature, correlation analysis between possible candidates and *near fiber count to mean distance ratio* were conducted. Having the highest correlation coefficient, *maximum slope*, was selected as the best index for representing significant MUPs. To ensure this correlation exists consistently for all MUs of different sizes, the correlation trajectory between *near fiber count to mean distance ratio* and *maximum slope* was performed relative to MU size (see Figure 4.6).

Assuming the slope as the best index feature, the correlation trajectories mentioned in Table 4.6 were then conducted.

Nandedkar et al. (Nandedkar et al. [1988c]) have reported a correlation coefficient of $r = 0.24$ between *number of fibers* in the MU and MUP *area* using a simulation model. Interestingly the correlation coefficient between MUP *area* and MU size including all MUPs in our dataset was the same. Moreover using our data, we could reveal the increase in this association when MUPs get more representative, i.e., maximum slope increases. It was shown that when MUPs with *slope* values lower than 0.6 v/s were excluded, the correlation coefficient between *area* and *number of fibers* in the MU was increased from $r = 0.24$ up to $r = 0.48$ and stayed constant thereafter, independent of *slope*. For relationships

between *duration* and MU size, the starting point correlation coefficient ($r = 0.31$) was the exact correlation coefficient mentioned between *duration* and *number of fibers* in the MU by Nandedkar et al. [1988c]. This value increased up to $r = 0.45$ when MUPs with slope values lower than 0.7 v/s were eliminated. Therefore, MUP *duration* was mentioned as a good representative of MU size (Stalberg and Karlsson [2001a]). Furthermore, simulation studies have suggested that *duration* measurements from only simple MUPs, contain information about the size of the MU (Nandedkar and Sanders [1989]). Most of the MUPs extracted from EMG signals acquired from normal muscles are simple shaped. Therefore it is expected to see high correlation coefficients between *duration* and MU size in healthy muscles as the case for the data presented in this thesis, since the muscle model simulated a normal muscle. Correlation analyses were conducted for *size index* and *thickness* and as a result the correlation trajectories between these two features and MU size were found to be similar to those presented for *area* and *duration* respectively.

The correlation trajectories performed between MU *average fiber density* and MUP features, all had a knee point at *slope* $= 0.6 - 0.7$ v/s except for *thickness* whose correlation was below 0.2 and independent of *slope*.

Size index and *area* of MUPs with *slopes* greater than 0.6 v/s, ($r = 0.3$) were determined to reflect *mean fiber diameter*. *Duration* was also a good determinant of *mean fiber diameter*, however to maximize relationships between *duration* and *mean fiber diameter*, MUPs should be further filtered (0.85 v/s). On the other hand, the correlation coefficient between *amplitude* and *mean fiber diameter* was pretty low ($r = 0.14$) but yet a knee point was observed. Neither *number of turns* ($r = 0.02$) nor *number of phases* ($r = -0.08$) were correlated with *mean fiber diameter* and the de-correlation was constant along the trajectory line.

Duration was shown to be not affected by *stdv of fiber diameters*. Based on the normality of the simulated muscle, most of the detected MUPs were simple. Therefore duration was affected by the size of the MU, rather than temporal dispersion caused by variation

in fiber diameters. Having said that, there was a slight increase in the correlation between MUP duration and stdv of fiber diameter when MUPs were filtered. This may be because the closeness of the fibers contributing to sharp MFPs to the MUPs, have exaggerated the effect of fiber diameter variation on duration.

On the other hand, *number of turns* was the only MUP feature characterizing the *stdv of fiber diameters*, where the correlation coefficient increased up to $r = 0.25$ for high slope MUPs. In contrast, *number of phases* was not related to *stdv of fiber diameters* and the maximum value for correlation coefficient was $r = 0.11$, even for highly significant MUPs. This explains why number of phases is not used for discrimination between myopathic and normal muscles. Variation in fiber diameter increases as an early sign in myopathic muscles, and since *number of phases*, is not sensitive to this factor, it can not discriminate between the diseased tissue and normal tissue.

All correlation coefficients at knee points along with corresponding slope thresholds for relationships existing between different MU features and MUP features are listed in Table 4.7. Based on correlation analysis, and as presented in this table, the low-slope-threshold was 0.6 v/s for most of the relationships. This threshold was increased to 0.7 v/s for relationships between *duration* and *number of fibers* in the MU, *thickness* and *number of fibers* and also *number of turns* and *stdv of fiber diameter*. For relationships between *duration* and *mean fiber diameter* the low-slope-threshold is the highest among other relationships, at 0.85 v/s.

Since VCF near fibers are related ($r = 0.94$) to those MU fibers in 500 μm of the needle electrode, the association between *near fiber count* and *MUP fiber count* was also assessed as a relationship between a MU characteristic and a MUP feature. The moderate correlation ($r = 0.45$) between these two features were further studied by corresponding scatter plot. According to the scatter plot, *near fiber counts* were higher than *MUP fiber counts*. This is explained by the fact that MFPs superpose in creating a MUP, hence suppress one another's sharp peak. Therefore the number of fibers contributing to MUP

sharp section, is underestimated in calculation of *MUP fiber count* and that's why the correlation coefficient is moderate and not high.

Pattern Discovery Analysis

Beside correlation analysis, a pattern discovery technique was applied to the same dataset trying to reveal inherent significant patterns. Two new statistics were introduced through which the discovered significant patterns were interpreted. The simple DCA and QCA statistics seemed to be useful enough to uncover the associations between MU characteristics and MUP features. These statistics were able to extract and demonstrate the information content of patterns discovered by PDT. Algorithms similar to the ones used for Sorted-Correlation were used and corresponding trajectories for these statistics were created.

To discover inherent patterns among MUP features and MU characteristics, the Sorted-PDT algorithm was applied to MUP features and MU features mentioned in Table 4.9, and corresponding trajectories were created. There was a knee point observed in all of the trajectories, meaning that MUPs with certain slope values are good representatives of their MUs.

QCA and DCA statistics suggested that residual values for discovered patterns between *area* and *number of fibers* in the MU, would be more significant if MUPs with slope values lower than 0.6 v/s and 0.4 v/s respectively were excluded. For the association between *duration* and MU size, DCA suggested that no filtering was needed. However, using the QCA statistic, MUPs with slope values lower than 0.37 v/s should be excluded, to maximize the relationship between *duration* and MU size. *Amplitude* was shown to be not a good feature characterizing MU size. Even so, the knee point at $slope = 0.42$ v/s was evident in the corresponding PDT trajectory. Trajectory between *thickness* and MU size was similar to *size index*, except the statistic values were lower.

Significant patterns were discovered between *average fiber density* and several MUP

features. Residual values discovered between *duration* and *average fiber density* were the most significant ones among other MUP features, thus *duration* was selected as the MUP feature being most affected by the *average fiber density*. Correspondingly, DCA suggested that MUPs with slope values lower than 0.6 v/s should be excluded in order to reach the maximum possible association between *duration* and *average fiber density*. However, the low-slope-threshold for the QCA statistic was a bit lower (0.47 v/s).

The Sorted-PDT algorithm was applied to *mean fiber diameter* and some MUP features. Among MUP features, DCA and QCA were the highest for patterns discovered between *duration* and *mean fiber diameter* and at the same time were constant independent of slope. Significant patterns between MUP *area* and *mean fiber diameter* were discovered as well as the ones between *size index* and *mean fiber diameter*. For *area* and *size index*, DCA statistic values suggested that MUPs with slopes lower than 0.5 v/s and 0.46 v/s should be excluded, whereas QCA suggested low-slope-thresholds to be equal to 0.7 v/s and 0.6 v/s.

The only MUP feature providing information about the *stdv of fiber diameters* in the MU is *number of turns* of the detected MUP. Nevertheless, DCA and QCA statistic values were not high and they dropped drastically after MUPs with slopes lower than 1.5 v/s were excluded.

The collection of knee points for DCA and QCA statistics have been summarized and mentioned in Table 4.10. The slope threshold at which the knee points happen were also listed.

Based on DCA statistics in Table 4.10, most of the relationships maximize for MUPs with slope above 0.5 v/s. For *duration* and *average fiber density* the threshold increases a bit to 0.6 v/s and to maximize *thickness* and *number of fibers* in the MU MUPs should be further filtered, $slope_{thr} = 0.7$ v/s. QCA suggests similar slope threshold of 0.5 v/s for most relationships. While, to maximize the relationship between *size index* and *mean fiber diameter* MUPs should be further filtered, i.e., $slope_{threshold} = 0.6$ v/s. As well for

relationships between *thickness* and MU size, and *area* and *mean fiber diameter* the slope threshold increased to be 0.7 v/s.

Comparing Tables 4.10 and 4.7, it was obvious that PDT results were similar to results gained from correlation analysis. Nevertheless, the low-slope-thresholds were lower for PDT analysis than for correlation analysis. In addition, PDT analysis suggested a high-slope-threshold as well as the low-slope-threshold suggested by correlation analysis. Based on trajectory plots for PDT analysis shown in Figures 4.18-4.21, there is a decrease in DCA and QCA for all PDT trajectories for high slope MUPs (MUPs with slopes ¿ 2.0 v/s). This means that high slope MUPs are not as informative and it might be better to exclude them from studies. Therefore PDT suggests that MUPs should be filtered based on slopes at both ends. This is not suggested by correlation analysis, for whatever reason, but definitely was obviously observed in PDT trajectories.

Chapter 6

Conclusions

To the best of the author's knowledge, this work was the first attempt to investigate relationships between MU and MUP features extracted from a complex physiologically based simulation model. The simulated data used in this thesis, was well resembling real data as reported in the literature. As well, the simulation model facilitated similar acquisition environments as the ones utilized by physicians in real clinical settings. Moreover, the decomposition techniques used in this research worked properly and generated similar results to what has been presented in the literature, using simple restricted models.

Having said all of the above, we are confident to report the following major outcomes of this research:

- The appropriate low-slope-threshold was suggested to be 0.6 v/s. Therefore, to maximize the relationships between MUP features and MU characteristics, only MUPs with slopes above 0.6 v/s should be considered.

- PDT results for the most part corroborated correlation analysis results. However, a high-slope-threshold not observed in correlation results, was recommended. PDT results have shown that high slope MUPs are not as informative about the underlying MU and could be excluded to maximize the relationship between MUP features and

MU characteristics.

- Some MUP features reflect specific MU characteristics best:
  - MUP *area* is the best representative feature for the size of the MU.
  - MUP *duration* is the best MUP feature to reflect *average fiber density*.
  - MUP *duration* was also a good representative for *mean fiber diameter*, however this was expected based on specific assumptions within the simulation model.
  - *number of turns* of the MUP reflects the *stdv of fiber diameter* in the MU from which it is recorded.
  - *MUP fiber count* was shown to be a good representative of MU fibers in 500 μm proximity of the needle electrode.

Some recommendations might be useful to build the road for possible future studies: 1) The problem studied in this thesis, is actually a multidimensional problem. Hence, it might be useful if the associations between MU characteristics, MUP features and volume conductor factors are investigated using multivariate analysis, such as multivariate regression. The pattern discovery technique could also be used, through which higher order patterns are investigated, however a proper visualization or interpretation method needs to be developed to make useful inferencing out of the high order discovered patterns. 2) It is believed that the associations between the features are not the same for normal muscles compared to diseased muscles. Therefore it is worthwhile to repeat the same research as presented in this thesis for myopathic as well as neuropathic data. 3) Given the different associations among MUP features and MU characteristics for normal and diseased muscles, it would be interesting to use them for real EMG signals and try to understand and predict the status of the underlying MU from associated MUP features.

Appendix A

Parameter Estimation

Given a set of data D, in order to discover the significant patterns, there are two variables to be estimated for each possible event - the expected occurrence of the event and the variance of the difference between the observed and expected occurrence of that event. To estimate both parameters, a default model has to be chosen first according to the problem domain and the available knowledge. Since *a priori* knowledge about the domain of the dataset is unknown, the model in which random variables are independent is selected. By this assumption the expected occurrence of a compound event would be defined as in Equation 3.3 while the variance, $c_{x_j^s}$, of the residual $d_{x_j^s}$ is given by:

$$c_{x_j^s} = e_{x_j^s}\left(1 - \sum_{x_{ip}\in x_j^s} \prod_{x_{jq}\in x_j^s, x_{jq}\neq x_{ip}} P(x_{jq}) + (|s|-1)\prod_{x_{ip}\in x_j^s}\frac{o_{x_j^s}}{M}\right) \tag{A.1}$$

$$e_{x_j^s}\left(1 - \sum_{x_{ip}\in x_j^s} \prod_{x_{jq}\in x_j^s, x_{jq}\neq x_{ip}} P(x_{jq}) + (|s|-1)\frac{e_{x_j^s}}{M}\right) \tag{A.2}$$

For details about the source of this formula, please see Wang [1997].

Bibliography

P.E. Barkhaus and S.D. Nandedkar. On the selection of concentric needle electromyogram motor unit action potentials: Is the rise time too restrictive? *Muscle & Nerve*, 19: 1554–1560, 1996.

P.E. Barkhaus, S.D. Nandedkar, and D.B. Sanders. Quantitative EMG in inflammatory myopathy. *Muscle & Nerve*, 13:247–253, 1990.

J.V. Basmajian and C.J. De Luca. *Muscles Alive; Their functions revealed by electromyography.* Williams & Wilkins, USA, 5 edition, 1985.

T.E. Bertorini, E. Stalberg, C.P. Yuson, and W.K. Engel. Single fiber electromyography in neuromuscular disorders: Correlation of muscle histochemistry, single- fiber electromyography, and clinical findings. *Muscle & Nerve*, 17:345–353, 1994.

Jaap Bethlem. *Myopathies.* Elsevier/North-Holland Biomedical Press, Amesterdam, 2 edition, 1980.

M.H. Brooke and W.K. Engel. The histographic analysis of human muscle biopsies with regard to fiber types: 1. adult male and female. *Neurology*, 19:221–233, March 1969a.

M.H. Brooke and W.K. Engel. The histographic analysis of human muscle biopsies with regard to fiber types: 2. diseases of upper and lower motor neuron. *Neurology*, 19: 378–393, April 1969b.

M.H. Brooke and W.K. Engel. The histographic analysis of human muscle biopsies with regard to fiber types: 3. myotonias, myasthenia gravis, and hypokalemic periodic paralysis. *Neurology*, 19:469–477, May 1969c.

F. Buchthal and P. Pinelli. Analysis of muscle action potentials as a diagnostic aid in neuromuscular disease. *Acta Medica Scandinavia*, 266 Suppl 142:315–327, 1952.

F. Buchthal, C. Guld, and P. Rosefalc. Action potential parameters in normal human muscle and their dependence on physical variables. *Acta Physiology Scandinavia*, 32, 1954.

B. Cengiz, F. Ozdag, U.H. Ulas, Z. Odabasi, and O. Vural. Discriminant analysis of various concentric needle EMG and macro-EMG parameters in detecting myopathic abnormality. *Clinical Neurophysiology*, 113:1423–1428, 2002.

Tom Chau. Marginal maximum entropy partitioning yields asymptotically consistent probability density functions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(4):414–417, April 2001.

D. Cros, P. Harnden, J.F. Pellissier, and G. Serratrice. Muscle hypertrophy in duchenne muscular dystrophy. *Journal of Neurology*, 236:43–47, 1989.

C. Doriguzzi, T. Mongini, L. Palmucci, E. Gagnor, and D. Schiffer. Quantitative analysis of quadriceps muscle biopsy. *Journal of Neurological Sciences*, 66:319–326, 1984.

Jacques Duchene and Jean-Yves Hogrel. A model of EMG generation. *IEEE Transactions on Biomedical Engineering*, 47(2):192–201, February 2000.

D. Dumitru and J.C. King. Motor unit action potential duration and muscle length. *Muscle & Nerve*, 22:1188–1195, 1999.

D. Dumitru, J.C. King, and S.D. Nandedkar. Motor unit action potentials recorded with concentric electrodes: Physiological implications. *Electroencephalography and Clinical Neurophysiology*, 105:333–339, 1997.

D. Dumitru, J.C. King, and W.E. Rogers. Motor unit action potential components and physiological duration. *Muscle & Nerve*, 22:733–741, 1999.

B. Falck, E. Stalberg, and C. Bischoff. Influence of recording site within the muscle on motor unit potentials. *Muscle & Nerve*, 18:1385–1389, 1995.

A. Hamilton-Wright and D.W. Stashuk. Physiologically based simulation of clinical EMG signals. *IEEE Transactions on Biomedical Engineering*, 52(2):171–183, February 2005.

Andrew Hamilton-Wright, D.W. Stashuk, and T Doherty. Simulatin jitter using a muscle model. In *AAEM annual meeting*, October 2002. Poster.

M.A. Johnson, G. Sideri, D. Weightman, and D. Appleton. A comparison of fiber size, fiber type constitution and spatial fiber type distribution in normal human muscle and in muscle from cases of spinal muscular atrophy and from other neuromuscular disorders. *Journal of Neurological Sciences*, 20:345–361, 1973.

S.D. Nandedkar. Models and simulations in electromyography. *Muscle & Nerve Supplement*, 11:S46–S54, 2002.

S.D. Nandedkar and D.B. Sanders. Simulation of myopathic motor unit action potentials. *Muscle & Nerve*, 12:197–202, March 1989.

S.D. Nandedkar, P.E. Barkhaus, D.B. Sanders, and E. Stalberg. Analysis of amplitude and area of concentric needle EMG motor unit action potentials. *Electroencephalography and Clinical Neurophysiology*, 69:561–567, 1988a.

S.D. Nandedkar, D.B. Sanders, and E.V. Stalberg. EMG of reinnervated motor units: A simulation study. *Electroencephalography and Clinical Neurophysiology*, 70:177–184, 1988b.

S.D. Nandedkar, D.B. Sanders, E.V. Stalberg, and S. Andreassen. Simulation of concentric needle EMG motor unit action potentials. *Muscle & Nerve*, 11:151–159, February 1988c.

C.S. Pattichis and A.G. Elia. Autoregressive and cepstral analyses of motor unit action potentials. *Medical Engineering & Physics*, 21:405–419, 1999.

G. Pfeiffer and K. Kunze. Turn and phase counts of individual motor unit potentials: Correlation and reliability. *Electroencephalography and Clinical Neurophysiology*, 85: 161–165, 1992.

B.G.H. Schoser, C. Schneider-Gold, W. Kress, H.H. Goebel, and P. Reilich. Muscle pathology in 57 patients with myotonic dystrophy type 2. *Muscle & Nerve*, 29:275–281, 2004.

M. Sonoo. New attempts to quantify concentric needle electromyography. *Muscle & Nerve*, 11:S98–S102, 2002.

M. Sonoo and E.V. Stalberg. The ability of MUP parameters to discriminate between normal and neurogenic mups in concentric EMG: Analysis of the MUP.

E. Stalberg and L. Karlsson. Simulation of EMG in pathological situations. *Clinical Neurophysiology*, 112:869–878, 2001a.

E. Stalberg and L. Karlsson. Simulation of the normal concentric needle electromyogram by using a muscle model. *Clinical Neurophysiology*, 112:464–471, 2001b.

E. Stalberg, B. Falck, M. Sonoo, S. Stalberg, and M. strm. Multi-MUP EMG analysis - a two year experience in daily clinical work. *Electroencephalography and Clinical Neurophysiology*, 97:145–154, 1995.

D.W. Stashuk. Decomposition and quantitative analysis of clinical electromyographic signals. *Medical Engineering & Physics*, 21:389–404, 1999a.

D.W. Stashuk. Simulation of electromyographic signals. *Journal of Electromyography and Kinesiology*, 3(3):157–173, 1993.

D.W. Stashuk. EMG signal decomposition: How can it be accomplished and used? *Journal of Electromyography and Kinesiology*, 11:151–173, 2001.

D.W. Stashuk. Detecting single fiber contributions to motor unit action potentials. *Muscle & Nerve*, 22:218–229, 1999b.

D.W. Stashuk and W.F. Brown. *Neuromuscular function and disease: Basic, clinical and diagnostic aspects*, volume 1, chapter Peripheral motor system: Quantitative electromyography, pages 311–348. Elsevier Science, Pennsylvania, USA, 2002.

I. Takehara, J. Chu, T.C. Li, and I. Schwartz. Reliability of quantitative motor unit action potential parameters. *Muscle & Nerve*, 30:111–113, 2004.

H. Tohgi, A. Kawamorita, K. Utsugisawa, M. Yamagata, and M. Sano. Muscle histopathology in myotonic dystrophy in relation to age and muscular weakness. *Muscle & Nerve*, 17:1037–1043, 1994.

C. Tosi and F. Jerusalem. Selektive muskelfasertypanomalien bei neuromuskularen erkankungen. *Journal of Neurology*, 214:13–34, 1976.

A. Uncini, D.J. Lange, R.E. Lovelace, M. Solomon, and A.P. Hays. Long-duration polyphasic motor unit potentials in myopathies: A quantitative study with pathological correlation. *Muscle & Nerve*, 13:263–267, 1990.

A. Verma, W.G. Bradley, N.W. Soule, W.W. Pendlebury, J. Kelly, and L.S. Adelman. Quantitative morphometric study of muscle in inclusion body myositis. *Journal of Neurological Sciences*, 112:192–198, 1992.

J.F. Wang, J. Forst, S. Schroder, and J.M. Schroder. Correlation of muscle fiber type measurements with clinical and molecular genetic data in duchenne muscular dystrophy. *Neuromuscular Disorders*, 9:150–158, 1999.

Yang Wang. *High-Order pattern discovery and analysis of discrete-valued data sets*. PhD thesis, University of Waterloo, Waterloo, Ontario, Canada, 1997. Systems Design Engineering.

Andrew K. C. Wong and Yang Wang. High-order pattern discovery from discrete-valued data. *IEEE Transactions on Knowledge and Data Engineering*, 9(6):877–893, 1997.

Andrew K.C. Wong and Yang Wang. Discovery of high order patterns. In *Intelligent Systems for the 21st Century*, volume 2 of *IEEE International Conference on Systems, Man and Cybernetics*, pages 1142–1147, Vancouver, BC, Canada, 1995.

Ewa Zalewska and I. Hausmanowa-Petrusewicz. Global and detailed features of motor unit potential in myogenic and neurogenic disorders. *Medical Engineering & Physics*, 21:421–429, 1999.

Ewa Zalewska and Irena Hausmanowa-Petrusewicz. Effectiveness of motor unit potentials classification using various parameters and indexes. *Clinical Neurophysiology*, 111:1380–1387, 2000.

Ewa Zalewska and Irena Hausmanowa-Petrusewicz. Evaluation of MUP shape irregularity - a new concept of quantification. *IEEE Transactions on Biomedical Engineering*, 42(6), June 1995.