Effective Information Display and Interface Design for Decomposition-based Quantitative Electromyography

by

Anne K. G. Murphy

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

Waterloo, Ontario, Canada, 2002

I hereby declare that I am the sole author of this thesis.
I authorize the University of Waterloo to lend this thesis to other institutions or individuals for the purpose of scholarly research.

Anne K. G. Murphy

I further authorize the University of Waterloo to reproduce this thesis by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

Anne K. G. Murphy

Borrower's Page

The University of Waterloo requires the signatures of all persons using or photocopying this thesis. Please sign below, and give address and date.

Abstract

Physicians perform qualitative assessment of electromyographic (EMG) studies to support diagnosis of neuromuscular disease. Quantitative analysis is not widely used. Decomposition-based Quantitative Electromyography (DQEMG) provides the ability to evaluate individual motor unit signals and statistics at high contraction levels, where typical EMG patterns are confusing. This study analyzes producing and presenting DQEMG signals for improved clinical utility.

Human factors research supported a prototype information display, which was evaluated by clinical experts and non-physicians for rapid collection, integration and comprehension of useful indicators of disease conditions. The expert users evaluated the display in the context of the DQEMG interface, the usability of which was also examined. Non-experts participated in a display mode comparison (text, histogram, polar star plot), which evaluated the displays by performance measures of error rate and speed of comprehension. The polar star plot representation was preferred by all physicians and the majority (81%) of non-physicians, providing intrinsic normative context and rapid assessment of signal characteristics. It produced the lowest error rates and was interpreted most quickly. A lack of workflow indicators and other non-optimal characteristics of the DQEMG interface were identified, with design suggestions offered for improvement. An integrated DQEMG information display that includes text reporting, histograms and a 6-dimensional polar star plot is recommended.

Acknowledgements

I must first acknowledge the support of my supervisor, Dr. Daniel Stashuk. I thank Dr. Stashuk for his guidance and insight regarding the field of decomposition-based quantitative electromyography. He has assisted me greatly in making it possible to do expert user testing and to implement some of my interface designs in the DQEMG application. The projects described here were also funded through NSERC funds allocated to Dr. Stashuk.

I also extend my gratitude to both the expert and non-expert users who participated in this study for the investment of their time and for the contribution of their insights. Thanks especially to Dr. Tim Doherty for lending his expertise as an electrodiagnostic physician to our research endeavors and helping us design the Expert testing scenario and questions in such a way as to clearly communicate with physicians.

I would also like to thank Dr. Catherine Burns and Dr. James Kay for their support and guidance in the research methods for this project. And I am pleased to thank my readers, Dr. Carolyn MacGregor and Dr. Catherine Burns, for their careful and critical review of this work.

A special gratitude must be extended also to a colleague overseas, Ewa Zalewska, whose support (and data) enabled me to complete the Nonexpert testing to my satisfaction.

Finally, I thank my family: William Murphy, who supported me first as a fiancée and then as a husband through the long process of qualifying for and then completing this degree; my parents, Eric and Diane, and my sister Sarah. I can never sufficiently express my appreciation for your love and encouragement.

Thank you.

All errors in this manuscript are strictly my own responsibility.

Table of Contents

List of Tables

List of Figures

List of Acronyms

ANOVA	ANalysis Of VAriance
AAR	Area-to-Amplitude Ratio
DQEMG	Decomposition-based Quantitative Electromyography
EMG	Electromyography
FR/MU	Firing Rate per Motor Unit
IDI	Inter-Discharge Interval
MUAP	Motor Unit Action Potential
MUAPT	Motor Unit Action Potential Train
MU	Motor Unit
MVC	Maximal Voluntary Contraction
NCS	Nerve Conduction Study
QEMG	Quantitative EMG

Chapter 1 Introduction

1.1. Overview

Electrodiagnosis is the use of electronically gathered information to assist in the diagnosis of neuromuscular diseases, such as muscular dystrophy. Since 1666 when Francesco Redi first deduced that muscles generated electricity, the medical profession has been improving its understanding of how, when, and why muscles produce these electrical signals, and how examining them can give us information about the condition of the muscle or, more recently, the neuromuscular system (Basmajian and DeLuca, 1985). The neuromuscular system includes the nerves that bring commands to muscle fibers (called motor neurons), the muscle fibers themselves, and the junctions between them whereby such commands are communicated.

The research for this thesis studied a larger system: the diagnostic system. The diagnostic system includes the neuromuscular system being examined (the patient), other information regarding the patient's case (symptoms, history), the electrodiagnostic examination and equipment (electromyography), and the physician whose diagnosis follows from the rest. This system exists in an environment that includes the medical profession in general and, more specifically, the training and experience of the individual physician. In current practice, electrodiagnosis is almost an art. The physician observes the EMG signals amid a sea of information and comes to some conclusion by comparing what is seen and heard to what is known. Excluding conduction velocity studies, which determine the speed of signal conduction, nothing in most clinical electrodiagnosis studies is measured quantitatively.

In the early 1950s, Fritz Buchthal introduced standards of measurement for a new type of *quantified* electromyography. He defined interesting characteristics of parts of the EMG signal, and published reference (normal) values for these characteristics in various muscles. But his technique took too much time to apply and people questioned how it contributed to the specificity of diagnosis. Fifty years later, and still for reasons of time limits and dubious contribution, quantitative EMG is not commonly used in clinical practice.

In the Biological Signal Detection and Analysis Lab at the University of Waterloo, Dr. Dan Stashuk and his students are working on computer applications that do something more sophisticated than the measuring by hand that Dr. Buchthal proposed. Yet these computer-mediated techniques are still quantitative EMG and continue to suffer drawbacks of time costs and unproven contribution to clinical decision-making. More details about DQEMG and the physiological basis of the signals it collects will be presented in the next chapter. In the years since

Dr. Buchthal first started to publish his work, many new characteristics of the EMG signal have been examined and adopted by the research field in quantitative EMG.

The research described in this thesis was an attempt to prove through analysis and performance measures how some EMG signal characteristics can improve the specificity and consistency of EMG interpretation. It also aimed to identify some productive avenues for reducing the time requirements of collecting and processing quantitative EMG data. The analysis included a system study and task analysis as well as a literature review. The experimental procedure included testing of expert users in their whole interaction with the research application, as well as testing nonexpert users on their ability to correctly interpret three different modes of displaying EMG signal characteristics.

1.2. Focus of Investigation

1.2.1. Effective Information Display

Attempting to ascertain a useful definition of "effective", we have analyzed several options for information display for the Decomposition-Based Quantitative Electromyography (DQEMG) application. Effectiveness in an information display is affected by two major factors: what information is chosen for display, and how that information is displayed to the user. Choice of the former was informed by the literature review on what characteristics of an EMG signal are useful in the diagnosis and monitoring of neuromuscular diseases, as described in Chapter 3. Some suggestions on how to present that information also arose from that review, but other suggestions evolved out of other information display work and theoretical understandings from cognitive ergonomics and ecological perception theory.

A muscle is composed of functional units called motor units. When a motor unit is active, a motor unit action potential (MUAP) can be detected from it. Neuromuscular disease affects the structure of the motor unit in a way that is reflected in characteristics of the MUAP. Many MUAPs add together to make the EMG "interference pattern", which is subsequently decomposed by DQEMG in order to identify individual MUAP trains (MUAPTs). In addition to characteristics of an averaged template MUAP for each train, DQEMG is able to report on the firing statistics of the whole train, as well as the percent of the patient's maximal voluntary contraction (MVC) that is represented by the contraction under examination. The standard list of MUAP characteristics is amplitude (peak-to-peak voltage), duration, area, turns, and phases. More recently characteristics have been combined into "indexes" that are considered more discerning: area-to-amplitude ratio

and size index (Sonoo and Stålberg, 1993). Other research asserts correlations between characteristics, such as duration, amplitude and %MVC, or turns and amplitude (Zalewska and Hausmanowa-Petrusewicz 1999, Buchthal, 1982). Some of the information that could be displayed follows a normal distribution within and between subjects, while some (mainly amplitude) do not.

The interest of this project was not to automate the classification of data into disease groups, but to present the DQEMG analysis in such a way as to improve the accuracy, efficiency, and overall effectiveness of the physician. Graphics designed according to ecological interface design principles provide a powerful means of display that makes the selected information quickly accessible to the busy physician, putting numbers in context and making distributions and outliers plainly visible. Nonexpert user testing was used to demonstrate that an ecological interface display, the polar plot, could be effective for DQEMG.

1.2.2. Time Factors and Other Barriers to Usability

A physician engaging in the study of a patient will review their history, give them a physical exam, look at the symptoms, and produce statements and ideas about the underlying physical condition of the patient. A physician doing an electrodiagnostic assessment will do all of these things and additionally observe nerve conduction velocities in the affected area as well as the form of EMG signals detected from various muscles in the patient. A technician may assist the physician in collecting the nerve conduction studies and preparing a report on the EMG study. Why? Because the physician's time is the most valuable resource in the lab.

Often a patient is referred into the neurology lab by another physician in order to eliminate a possible diagnosis as much or more than to produce one. Qualitative EMG is a sensitive method, but not a specific one. That is, an abnormal condition is easily distinguished from a normal condition, even if the abnormal condition is not drastically advanced; "a properly performed and interpreted needle EMG examination is rarely abnormal in normal subjects (AAEM 1999)." Identifying the specific type of abnormal condition is not so easy; currently needle EMG is not a specific measure. Our hypothetical clinical EMG physician conducts all of his examinations and analysis for a single study in 15 minutes to an hour. One EMG study might include just one muscle but will more likely survey multiple muscles to determine the course of a disease down a limb or through another part of the body.

Stålberg et al. (1995) described how in the 1980s computer-assisted automatic analysis brought the time requirement for an investigation of 20 Motor Units (MUs) in one muscle down to around 20 minutes. "This was too slow," they said, "to make even this technique widely accepted

for routine work (p 145)". After doing Multi-MUP analysis for two years clinically, in the '90s that research group had their technique down to 4-8 minutes. Multi-MUP analysis is similar to the decomposition-based method discussed here, but not exactly the same. DQEMG makes more of an attempt to identify all of the MUPs in the interference pattern. Doherty and Stashuk (1999) reported the time to collect and edit 20 MUAP trains with DQEMG was about 10 minutes. Considering that a typical EMG study might aim to characterize more than one muscle, one is left with an analysis technique that might take the physician as much time as all the rest of the examination and study. This is probably still too long for it to be commonly adapted.

In discussions with Dr. Doherty, (a colleague at the Dept. of Physical Medicine and Rehabilitation at The University of Western Ontario in London) we have identified some problem spots in the application. Though collecting each contraction (of which one might collect 3 or 4 to get 20 MUAP trains identified) only takes 30 seconds to a minute, editing the MUAP trains and characteristics to make sure they are all valid can take more than 10 minutes. Some of the time this takes may be due to poorly designed details of the interface. How the interface is currently designed is discussed in more detail in Chapter 4. Our analysis has identified points for improvement, and the Expert user testing has confirmed some of those and suggested others.

In general the usability of DQEMG needs to be understood given the context of the physicians' work; her goals and expectations as well as her skills and knowledge for manipulating a computer application. While we don't expect an untrained user to sit down and fly through the program on her first try, the program could be made more usable and transparent. Right now there is no suggested workflow in the interface, and no help or instructions. In the course of this project we have written a general introduction to DQEMG for the physician, and analyzed how workflow might be improved and made more obvious in the program. Commonly used screens can be made more obvious or accessible and an instructional context could be included on some screens. Also, we attempted to identify places where labels could be made more effective and considered whether or not a modal type of control was in keeping with the users' skills and mental models, as well as the task requirements.

Beyond the background information of chapters 2 through 4, the design and rationale for both the Expert and Nonexpert user testing protocols are presented in chapters 5 and 7. This includes a description of how normal, myopathic and neuropathic data was simulated for the Nonexpert testing. Experimental results are reported separately in chapters 6 and 8 and discussed together in chapter 9, which also covers limitations and points for further research. Concluding comments including a summary of DQEMG redesign recommendations are in chapter 10.

Chapter 2 Background

2.1. Electromyography and Muscle Structure

Electromyography is the science of detecting and interpreting EMG signals. The signals, which are detected using specific types of electrodes, reflect the physiology of the muscle involved. Electrodiagnosis is based on the idea that EMG signals follow patterns that reflect disease and other changes in neuromuscular physiology in identifiable ways. This is an introduction to the muscle structures and behaviors that most affect the EMG signal and the basic characteristics of that signal, as well as the equipment and applications used to report about them.

2.1.1. Motor Units

Muscles are made up of muscle fibers, which are long and strand-like and generally connected to bones at either end by tendons. In roughly the middle of each muscle fiber is a neuromuscular junction, also called the motor end-plate. This is where the fiber is innervated by an axonal branch of a motor neuron. One neuron innervates a group of muscle fibers that together are called a motor unit (MU). When an action potential, a self-propagating voltage wave, travels down that motor neuron it is transmitted to the muscle fibers of the motor unit via chemical reactions at their neuromuscular junctions.

Each MU consists of between 9 and several hundred muscle fibers, which are randomly distributed in a motor unit territory of approximately 2 to 10 mm diameter (Stålberg and Falck, 1997). The number of muscle fibers per MU varies from muscle to muscle; the biceps is estimated to have approximately 200 muscle fibers per MU (Stålberg and Falck, 1997). Muscle fibers belonging to different motor units are interspersed with one another. During a contraction, different motor neurons will fire, or transmit action potentials to motor units, at different times. All of the contractions of different motor units add together to achieve what is perceived to be a single smooth muscle contraction.

2.1.2. Muscle Fiber Action Potentials (MFAPs)

If the action potential is successfully transmitted across the neuromuscular junction a muscle fiber action potential (MFAP) propagates in a wave from the neuromuscular junction out toward the two ends of the fiber, causing it to contract. The muscle fiber contraction takes 50 ms to 200 ms depending on whether the type of muscle fiber is fast or slow twitch. The electrical activity in the

muscle fiber dies out in the wake of the action potential and can be stimulated again as little as 3 ms later. If the muscle fiber is still contracting, repeated stimulation can maintain the fiber in a constant contraction or tetanus state. The action potential triggering, or firing, as it is more commonly called, is an all-or-nothing event; it either happens or it doesn't. The time the MFAP takes to travels down the length of the muscle fiber depends on the length and diameter of the fiber. Diameters of muscle fibers vary from 10 to 100 μm. The MFAP can be detected by a needle electrode. This is the extracellular potential, measuring the volume conduction voltage around the muscle fibers. This is not a direct detection of the action potential across the muscle fiber membrane; the voltage of the detected potential will be smaller than the transmembrane potential.

2.1.3. Motor Unit Action Potentials (MUAPs)

A Motor Unit Action Potential, or MUAP, is a summated action potential as detected from all the muscle fibers in the same motor unit. It is the summation of all the MFAPs produced by fibers of the MU. The shape and characteristics of a MUAP are shown in Figure 2-1.

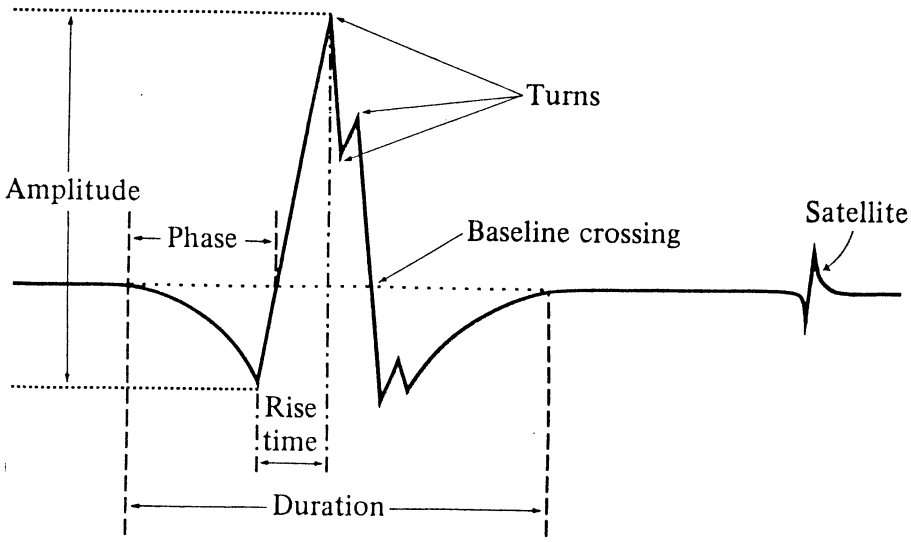


Figure 2-1 Characteristics of a MUAP (from Basmajian and Deluca, 1985)

The peak-to-peak voltage of a MUAP is called the amplitude. It is measured in microvolts (μV). The amplitude will depend on the distance from the electrode to the muscle fibers in its detection range. The duration of the MUAP is the length of time the MUAP can be distinguished from background noise, measured in milliseconds (ms).

A separate phase is counted for each time the path of the MUAP departs from 0 and returns, subject to a minimum amplitude threshold of 20 μV. Turns are deflections (places where the derivative of the waveform changes sign) that are at least 25μV in amplitude (Stashuk, 1999). Normally MUAPs have about 3 phases. If more than 10% of the MUAPs in a study are polyphasic, or have more than 4 phases, that study is considered abnormal, or diseased (Preston and Shapiro, 1998).

2.1.4. The Interference Pattern (EMG)

When a patient is maintaining a low level of muscle contraction, individual MUAPs are easily visible in a display of the EMG signal from a concentric needle electrode. As contraction intensity increases, however, more motor units are recruited and the firing times of the motor units get closer together. Different MUAPs will overlap, causing an interference pattern in which the human eye cannot consistently discern individual MUAP shapes or firing patterns.

2.1.5. Electrodes and Electromyographs

The three most common types of needle electrodes for electromyography are (in order of detection area size) monopolar, concentric, and single fiber. Surface electrodes are also used, both to collect surface EMG and to provide a reference for the needle EMG signal. Concentric needle electrodes are recommended for collecting contractions into the DQEMG application for clinical purposes. DQEMG is an analysis program, which depends on the use of an electromyograph to collect the EMG signal, sample it into a digital signal, and send the resulting data to DQEMG for analysis.

The Neuroscan Comperio system is the latest model electromyograph that is designed to output data into DQEMG. The Comperio System includes a Windows ™ -based PC, and keyboard and mouse, as well as a specialized operating board with a roller ball mouse and knobs for controlling such things as sensitivity, sweep, and volume of the speaker. The amplifier has input connectors for electrode and ground wires, as well as a speaker that allows the physician to listen to the EMG signal. Most of these are standard electromyograph features, though older models run on standalone machines, not PCs.

The Comperio signal acquisition program is called EMG/EP. The EMG/EP application is set up so that when the user hits a button called “Analyze” (or a function key), EMG/EP calls DQEMG to record and analyze a sample of the EMG signal. At that same time it passes patient and muscle information to DQEMG which the DQEMG application uses to name the folders in which it saves the data from that muscle study. DQEMG has an “Acquire” button that sends the

user back to EMG/EP in order to add further contractions to an existing study. In this way, the two applications are currently integrated together.

2.1.6. Decomposition-Based Electromyography (DQEMG)

The procedure for a quantitative electrodiagnostic study to characterize a particular muscle has three main steps. First the practitioner is required to perform an MVC protocol, where the patient performs a maximum voluntary contraction and a surface EMG signal is recorded into DQEMG. Then the practitioner samples the activity of at least 20 motor units at a moderate level of contraction with a concentric needle electrode. This typically takes 3 to 6 contractions but may require more depending on the condition of the muscle. After the acquisition of each contraction's EMG the DQEMG program decomposes the signal and isolates the activity of individual MUs. The interference patterns acquired during muscle contractions are thus decomposed into their constituent Motor Unit Potential Trains (MUAPTs). To do this, the

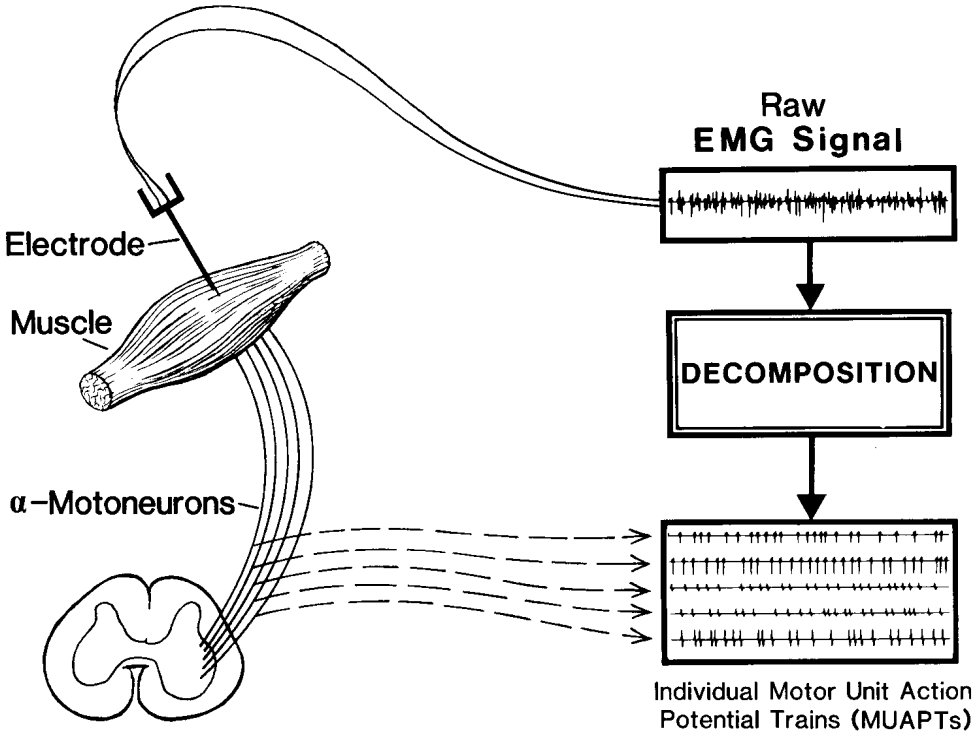


Figure 2-2 The decomposition transformation from needle–detected EMG signal input to individual MUAPTs (from Basmajian and Deluca, 1985).

DQEMG program identifies unique portions of the signal and classifies them to determine which MUAPs are created by which motor units. For each motor unit it then calculates a template MUAP to represent the average shape of the MUAPs created by that MU. Firing rate information is also used for classification, since a MU usually fires at pseudo-regular intervals. The interval between one firing and the next is called the inter-discharge interval (IDI). The final part of the quantitative electrodiagnostic study is when the physician reviews and assesses the results of the decomposition, both to confirm that the decomposition of each contraction is acceptable and the landmark positions on the MUAP templates are valid, and to analyze the character of the whole muscle study and consider its clinical implications.

2.2. Neuromuscular Disease

There are many kinds of neuromuscular disease, and this research does not attempt to consider the complexity of indications for each of them. However, most neuromuscular diseases fall into three main categories: neuronal diseases and conditions, also called neuropathies; myopathies; and diseases of the neuromuscular junction. Diagnostic indications for the latter type of disease will not be considered in this report.

2.2.1. Neuropathy

There are two types of neuropathy, axonal loss and pure demyelination. The second of these occurs when the myelin sheath that protects and supports the axon is damaged but the axon is not lost. Pure demyelination reduces the conduction speed of the axon but does not affect the morphology of the MUAP (Preston and Shapiro 1998), so it is not discussed here.

When an axon is lost or a motor neuron dies, the muscle fibers that were innervated by it either die or are gradually re-innervated by neighboring neurons. So motor units become fewer in neuropathy, but they are larger and stronger. In the early stages of neuropathy the patient presents little or no loss of muscle strength. As the disease condition progresses, the fine motor control of the patient will worsen as the patient runs out of small, low-recruitment-threshold motor units. Fewer motor units produce the same amount of force, so recruitment will be low. Eventually the patient will show loss of muscle strength. Denervation and reinnervation worsen the synchronicity of muscle fiber firing, causing the complexity of the MUAP to increase.

2.2.2. Myopathy

Myopathy is a disease condition in which the muscle fibers are lost or dysfunctional, rather than the axons. The patient retains the same number of motor units, but they are smaller, with fewer muscle fibers in each motor unit. Because they are smaller, more motor units may be recruited earlier in a contraction to achieve the same level of force. The myopathic patient will gradually lose muscle strength. Increased variability in the diameters of the muscle fibers leads to different conduction velocities and increased temporal dispersion, which will make the detected MUAP waveforms more complex.

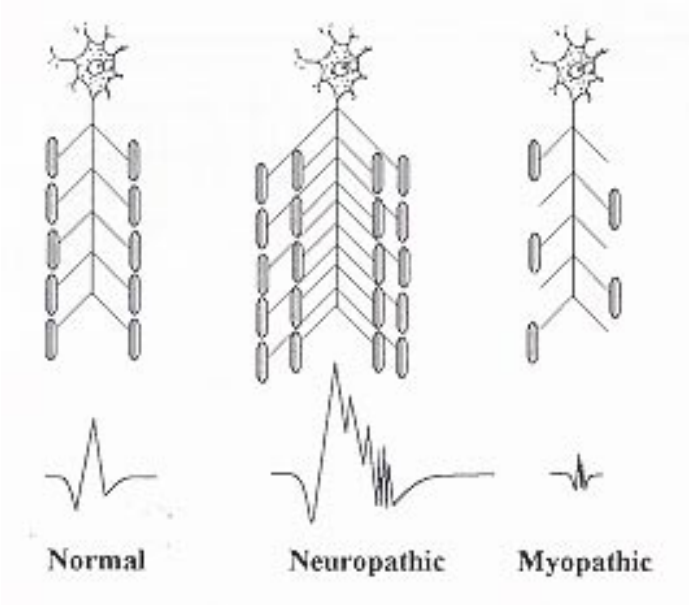


Figure 2-3 Physiology and MUAP morphology in normal, neuropathic and myopathic conditions. Reproduced from Preston and Shapiro (1998).

2.2.3. EMG Parameters in Health and Disease

Different muscles have different ranges of characteristic values in normal and disease cases, as do different patients. Despite 50 years' practice of quantitative EMG, reference values from normal subjects are still only available for a small selection of major muscles. The most comprehensive study of action potential parameters in different muscles continues to be the one published by Buchthal and Rosenfalck in 1955 (as cited by Preston and Shapiro, 1998). That report is organized by age since some action potential parameters vary with age. Other studies report variation with gender but that has not been conclusively proven.

The properties of MUAPs collected with concentric needle electrodes vary with the % maximal voluntary contraction (MVC) at which they are collected. For example, as the % MVC rises from threshold to 30% MVC in brachial biceps of normal subjects, amplitude mean and standard deviation (SD) can increase by as much as 60%. Duration mean and standard deviations, on the other hand, go down by almost 30%. The mean number of turns goes up but only slightly, and the SD stays the same. The firing rate mean also goes up by almost 60% but the firing rate SD does not change significantly (Howard et al 1988). This is due to changing recruitment at different levels of contraction, so that different motor units are detected, with different characteristics.

Generally speaking, neuropathies are characterized by large, polyphasic MUAPs with long durations and high amplitudes, though after chronic axonal loss the MUAPs may no longer be noticeably polyphasic. Chronic myopathy is identified with polyphasic MUAPs that are small and thin, with short duration and low amplitude. Some myopathies are observed to involve normal or even increased amplitudes and durations. Preston and Shapiro (1998) assert that the only reliable way to tell between neuropathy and myopathy is through examination of the recruitment pattern. However, besides unspecific terms like early, late, normal and reduced, they don't give much guidance to the practitioner as to how to recognize these patterns.

When a physician performs a needle EMG study, he samples the voltage potential from a few different locations in the muscle, hoping to sample the activity of a representative set of motor units. Myopathies and neuropathies may both have motor units that produce normal-looking MUAPs. It is not possible to characterize a muscle by looking at a single MUAP; the distribution of EMG characteristics has to be understood in order to characterize a muscle study and detect the physiological conditions that underlie the study results. This is why statistical and graphical methods of reporting information about that distribution are being used and investigated, and it is why physicians are advised to collect MUAPs from at least 20 MUs before they draw any conclusions. More specific patterns and ranges of useful EMG parameters in disease will be discussed in Chapter 3.

2.3. The Diagnostic Task

In order to design an interface and display for quantitative electrodiagnosis, we first conducted an analysis of the existing clinical diagnostic system. This system includes the user, or physician, and the assisting technician, the patient, the electromyograph and other equipment, the purpose and transformation inherent in the system, and the environment in which the system operates.

2.3.1. User Profile

The main user for this system is a professional physician. Other possible users include students and technicians, who would probably be using the system either to study and learn how to use it or in order to assist a physician. The primary users are trained in medicine, physiology and disease, and may have additional training in electromyography, neurology, or rehabilitation and physical medicine.

The physician user is most likely male but perhaps female, and could be middle aged or older. Students are most likely in their twenties, while technicians might be any adult age. Some users may have difficulty seeing due to advanced age or color blindness. They might tire easily, or have trouble maintaining information in short-term memory.

It is assumed that most of the users of this application will speak and read English. As professionals in a technical field, all of them can be expected to have some familiarity with computers, and they will know how to use an oscilloscope and an electromyograph and also how to interpret the information typically displayed by this equipment. They are also expected to be physically able to use such equipment, and to perform a needle exam on a patient.

The user is probably not an expert in installing and maintaining computer hardware or software. He might be a novice at that sort of thing. He probably has experience using Personal Computers (PCs) - a Macintosh or an IBM clone running some Microsoft OS, DOS, or possibly a centralized system made specifically for a hospital or school.

All of our potential users are pressed for time. This is their most precious and limited resource. In the clinic, the doctor's time is considered one of the most valuable resources in the system. Students and technicians will also be short on time. All of the potential users need a system that works in the least amount of time possible. Working in a stressed and rushed environment could increase their decision-making tunnel vision, making it harder for them to consider and pay attention to input that is unexpected or that does not confirm expectations (Wickens, 1996).

All of the users are expected to be familiar with the basic patterns of neuromuscular disease, both physiologically and in conventional EMG. They should be familiar with the definitions of Amplitude and Duration, though some typically use peak-to-peak amplitude while others more commonly consider negative peak amplitude, or the potential difference between the negative peak and the baseline. Similarly most will know what is meant by signal area but some might expect "Area" to be just the negative peak area. While many users will relate to the qualitative concept of signal thickness, not all of them will be familiar with the normal or disease values of

the amplitude-to-area-ratio. Relatively new indexes such as size index were probably not taught to most users in school; though they might have read about them in a research journal, they are not expected to have ever really used them.

2.3.2. System and Environment

Electrodiagnosis is based on the idea that EMG signals follow patterns that reflect altered physiology and therefore disease in identifiable ways. Physicians currently make mainly qualitative assessments of these patterns in the context of other information they have about disease and the patient being examined. The patients become an active part of the diagnostic system when they report their own symptoms and history. Information the physician observes and has learned is part of the environment of the system, along with whatever laws and professional standards govern the practice of electromyography and the technical resources that are available.

A Customer, Actor, Transformation, Weltanschaung, Owner, Environment (CATWOE) analysis is recommended for understanding human activity systems (Checkland, 1981). Some of this analysis was taken from an unpublished system study by the author under the name Anne Gay (2000).

- **C**ustomers: Patients, other Physicians
- **A**ctors: Physicians, Technicians, Patients
- **T**ransformation:

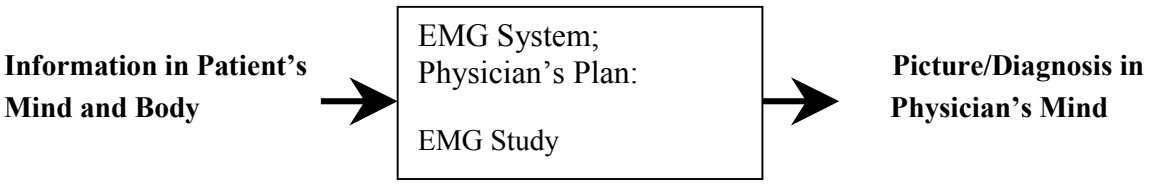


Figure 2-4 The diagnostic system transformation.

- **W**eltanschauung: EMG signals reflect physiology in patterns that are recognizable and consistently indicative of neuromuscular diseases. The gestalt form and sound of an EMG interference pattern, when interpreted by a skilled physician in light of other clinical information, can be an aid to diagnosis
- **O**wner: Medical Profession, Biomedical Engineers
- **E**nvironment:

  The National Economy

The Government (Taxes/Laws)

The Health System

The Educational System

Engineering & Technology (+Resources)

Medical Theories of Health and Disease

The patient's resources include time, money (health care/insurance), and self awareness. The patient's resource of trust in the physician might also affect his or her self-reporting. The physician's resources in this system are time, medical knowledge, attention, equipment and computer programming.

2.3.3. Clinical Procedure

The clinical procedure includes both the patient encounter and the reporting process for collecting and interpreting the EMG data. In preparation for the DQEMG interface design, an observational study of clinical protocol at University Hospital in London, Ontario was undertaken. Over the course of a day, multiple clinical examinations were observed. The equipment and electromyograph displays were videotaped in cases where the patient gave permission. Actual procedure was found to be different from textbook procedure.

Patient Encounter

1. Take history and perform directed physical examination
2. Formulate a differential diagnosis
3. Formulate a study based on the differential diagnosis
4. Explain test to patient
5. Perform nerve conduction studies
6. Perform needle EMG study

Figure 2-5 Procedure for a patient encounter (Preston and Shapiro, 1998)

In the London EMG lab the nerve conduction studies were performed by a technician and reviewed by the physician before the physician took the patient history. In many cases the patient had been referred by another physician, so some written history had been available prior to meeting with the patient. This written history guided the nerve conduction study design. The physician formulated a needle EMG study, explained the test to the patient, and performed the needle EMG study while reporting a set of results to the technician, who typed them into a

computer using an in-house report generating application. The physician then performed a directed physical examination in another room while the technician prepared the lab for the next patient. The physician could then go to his office to dictate his report into a tape recorder while the technician started nerve conduction studies on the next patient.

The subjective EMG report that followed from the clinical EMG study covered such things as Insertional Activity, Fibrillation Potentials, Positive Sharp Waves, Amplitude, Duration, Polyphasic activity, and Recruitment Pattern. These were reported using qualitative terms like Normal, Decreased, Increased, Occasional, Reduced, Very Reduced, Some, Many, Full, etc. Though a report could include indications of possible diagnoses, most reports eliminated diagnostic possibilities without drawing definite diagnostic conclusions. A report might also recommend further tests or treatment. Although the waveform data is kept on file in the lab, when a patient has a repeated test, they compare the qualitative reports (plus conduction velocity numbers) to one another, not the waveforms or waveform characteristics.

Errors in executing this procedure can include anatomical errors, errors in technique, or errors in interpretation. A 1976 study of 112 electromyographic reports found 86% to have substantial errors of interpretation (Johnson, Fallon, Wolfe, 1976).

2.3.4. Detailed Transformation

In light of the observational findings, a more detailed version of the diagnostic transformation was produced.

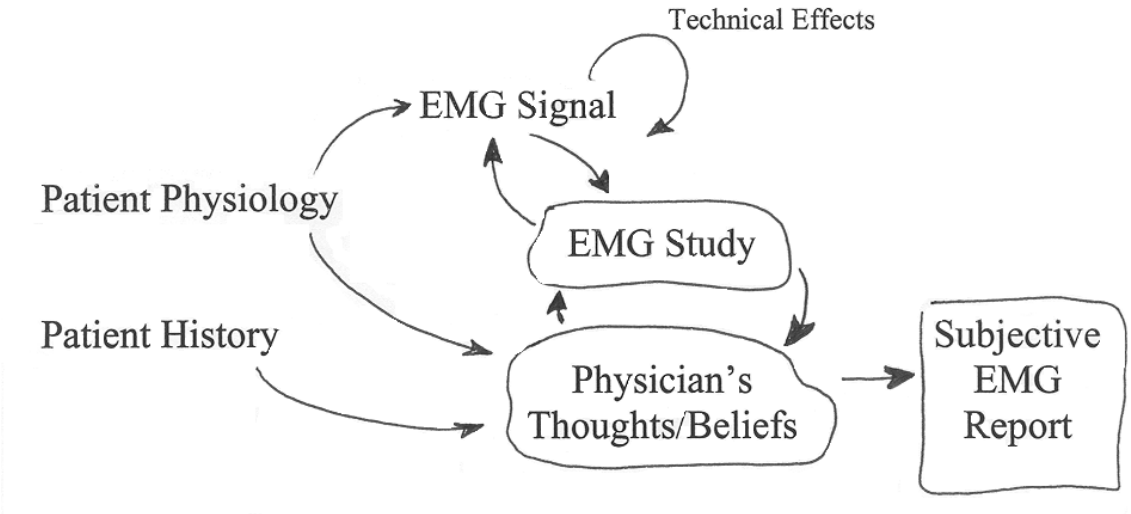


Figure 2-6 Detailed Electrodiagnostic Transformation

There is noise in the system caused by incorrect information and misinterpretation. The physician has to be careful to distinguish spontaneous activity in the muscle from technical effects such as insertional noise from the electrode, and noise in the wires and amplifier. The physician's thoughts and beliefs are affected by input from the EMG study and also lead to further EMG investigation, causing a cycle of feedback in the system. Though it is not shown in this diagram, the physician's beliefs about medical possibilities also guide the questions that are used to solicit the patient history.

2.4. Ecological Perception and Interface Design

According to the theory of ecological psychology and visual perception, human beings have an innate ability to perceive shapes and patterns, geometries, discontinuities, and symmetry (Gibson, 1986). An interface designer can take advantage of this direct perception of global features in a way that makes display interpretation and state recognition faster and easier, requiring less working memory and attention resources than a non-ecological display.

A classical example of an ecological interface design is one in which a set of hemispherical dials is arranged vertically, with all of the scales normalized so that the indicator rests in the middle of each dial so long as the system is running smoothly (in race cars, this is called having "clocked" dials). Under normal operating conditions, the operator can automatically recognize the unbroken vertical line of the indicator arrows lined up in this way. The vertical line created by the indicators is called an *emergent feature* since it emerges from the coordinated positions of multiple indicators. When the system is in an abnormal state the vertical line is broken, and the fact that something is wrong can be directly perceived by the operator, with little time or mental resources spent on interpreting the display.

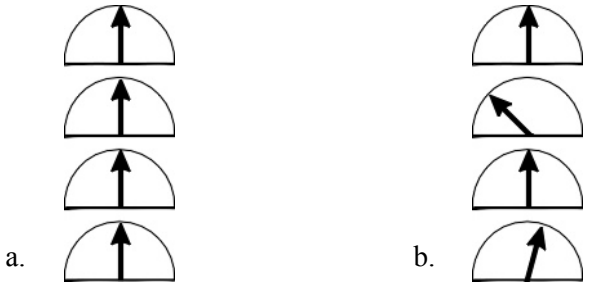


Figure 2-7 A clocked display in a) normal and b) abnormal conditions.

Part of ecological interface design, then, involves designing emergent features into the display whenever it is possible to do this in a meaningful way. In a situation like EMG where there are multiple system parameters to be interpreted, more than one variable can be combined

into the same graphic or neighboring graphics to produce an emergent global feature such as a polygon. Various geometric characteristics can be used meaningfully in this case, such as size, shape, angle and symmetry (or asymmetry). An information display that uses geometrical shapes aims to take advantage of the user's natural ability to perceive and recognize shapes and patterns. One type of display like this is called a polar star display (examples and more detail are given in Chapters 4 and 7). Polar-star displays were developed in the aviation industry in order to help users detect deviations from normal states based on valuable groupings of parameters. Some researchers in that area have also theorized that this type of display would be useful for monitoring patient health in the medical industry (Trujillo and Schutte, 1999)

Ecological interface design is based on an understanding of the underlying physical laws and principles involved in the task. So, following a system study and analysis of the diagnostic task, a literature review of both clinical diagnostic literature and pattern recognition literature was completed in order to understand the physical basis of MUAP characteristics and the patterns they follow.

Chapter 3 Analysis of EMG Characteristics

This chapter reviews the EMG characteristics that could be presented in an information display, and their distributions and patterns in diseased and healthy muscle, and includes some discussion about displaying them graphically. In a review of EMG in clinical diagnosis and research the same characteristics were generally used by clinical physicians as by computerized pattern recognition systems under development. The most common characteristics by far are Duration and Amplitude. Some researchers have proposed mathematical indices or coefficients that combine other characteristics into a single number that is seen to represent an aspect of the MUAP such as size or irregularity. This chapter analyses the characteristics in terms of their current use, their potential usefulness to the physician, and properties of them that lead to guidelines for displaying information about them. It will cover the following characteristics one may report on:

Duration
Amplitude
Area
Area/Amplitude, the Area to Amplitude Ratio (AAR) [Thickness]
Size Index
Firing Rate
Firing Rate per Motor Unit (FR/MU)
# of Phases
% Polyphasic MUAPs
# of Turns

Firing rate is a characteristic of the MUAP train, not just a single action potential. But all of these characteristics could be used to describe an individual motor unit. Presenting the statistical average or mean of a MUAP characteristic from a sample of MUs may provide an effective way to characterize the whole muscle. In some cases such a statistical summary seems effective, but in others it does not make as much sense. In particular, ways of identifying outliers should be retained.

Though in many cases the distributions of MUAP characteristics in particular disease cases across the entire population are well established or described, the distributions of these characteristics within a single individual have not been described in the literature. While recognizing the limitations of this approach, we have generally assumed that individuals within a

disease group present roughly the same distribution of MUAP characteristics within themselves as are observed in the whole group.

3.1. Duration

Duration is defined as the time from initial deflection from baseline noise to the final return of the MUAP to baseline. It reflects the number of muscle fibers within a motor unit (Preston, Shapiro, 1998) as well as the overall motor unit territory and parts of the muscle fiber physiology which affect timing, such as conduction velocity and motor end-plate locations. This is a quantitative characteristic that can also be qualitatively evaluated. It can be visually evaluated at low-level muscle contractions and by sound (frequency) even at mid-level contractions. Overly long duration MUAPs give a low thudding sound and very short MUAPs produce a high-frequency, scratchy sound. Duration is a valuable characteristic of EMG both because it changes predictably with physiology and disease and because it does not change drastically with the distance of the EMG needle electrode from the Motor Unit.

The DQEMG program is reasonably accurate at identifying the onset of a MUAP but can be less accurate at identifying the offset point, or end, of a MUAP. The adoption of both these points "can be affected by noise, baseline fluctuation, and other artifacts" (Zalewska and Hausmanowa-Petrusewicz, 1999). So long as a skilled user evaluates and corrects the identification of these landmarks, this is a fairly precise characteristic. Normal duration ranges in a variety of muscles are well known, as well as the effects on those ranges of age and temperature.

Preliminary investigations indicate that duration distribution within a single muscle on a single subject at one time is fairly Normal, or Gaussian. Observation of a histogram of Duration values within a patient may be useful, but this lab has not yet encountered a patient whose duration histogram did not center in a close and symmetrical fashion around the study mean. On the other hand, a visual display of the distribution could raise the confidence of the operator in the results reported.

The distribution of duration as a characteristic of individuals with specific disease types is widely distributed, with different diagnostic classes having distributions that overlap considerably, mostly due to large standard deviations. The normal range of duration is 5 to 15 ms, with a maximum around 20 ms (Preston and Shapiro, 1998). Neuropathic individuals are more likely to present with a longer duration than normal, and Myopathic individuals are more likely to present with reduced duration, with respective ranges of 4 to 20 ms (neuropathy) and 1 to 15 ms

(myopathy). This means that there is no strong argument for providing a visual indicator on a duration histogram that suggests a likely classification of a muscle study.

On the other hand, Buchthal and Pinelli described in 1956 the distributions of duration in normal and myopathic patients (see Figure) and they look very different. The Myopathic patients have both many short duration MUAPs and a few mid duration MUAPs, making their distribution uneven to the low end. If this is consistent in quantitative EMG, that supports the arguments that duration should be displayed in a distribution graph or histogram.

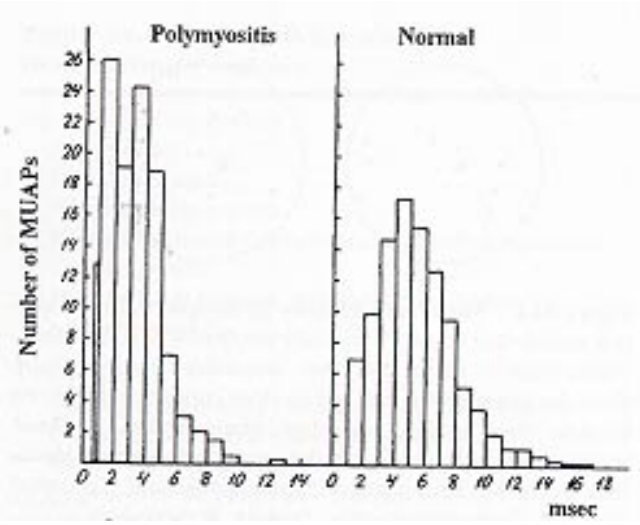


Figure 3-1 Histograms of MUAP duration in normal and myopathic patients (Buchthal and Pinelli, 1956, as cited in Preston and Shapiro 1998)

Normative values vary significantly by muscle. Stashuk and Doherty (unpublished) report Duration in the First Dorsal Interoseus to have a mean of 9.2 ms ± 1.9 in normal controls, with a range of 5.8 – 11.7 ms. The low end of this range is well below the low end of the normal range for Biceps, which they report as 8.3 – 12.8 ms with a mean of 10.8 ms ± 1.5. The Biceps ranges are similar to those reported Buchthal and Rosenfalk in 1955 for subjects aged 30-49 (9.0 – 12.1 ms) (As quoted in Preston and Shapiro, 1998). Durations are known to get longer with age, so presumably Stashuk and Doherty had subjects in the 25-49-year age range.

3.2. Amplitude

Amplitude is a more variable characteristic than duration, since it depends strongly on the position of the needle electrode. DQEMG reports the peak to peak voltage, which is a standard measure, though some other labs use the voltage difference between the largest negative peak and the baseline for EMG studies. In clinical conditions physicians currently assess amplitude

qualitatively by viewing it on an oscilloscope-like display and comparing the height of the signal with other signals in their memory. They can do this partly because they usually maintain the same visual sensitivity, or vertical scale, on their monitor.

Amplitude has a Log Normal distribution within an individual, the range of which varies with disease classification. Zalewska and Hausmanowa-Petrusewicz (2000) found that Amplitude is consistently reduced or normal (5 to 850 μV) in myopathic cases, but can vary between reduced and very high for neuropathic cases. Still, the maximum amplitude observed in neuropathies (>3000 μV) is outside the normal range for amplitude, which is 100 to 2000 μV. Using log(amplitude) will produce a more normal distribution in an amplitude data display (Stålberg, et al, 1996). Another option is to display amplitude on an axis with an exponential scale.

It may be valuable to display the range and distribution of amplitude to the physician, since atypical amplitudes (that is, MUAPs with amplitudes that are atypical for the state of the muscle) are not uncommon (Zalewska and Hausmanowa-Petrusewicz 2000) and can skew statistics such as mean and standard deviation easily. Placing this display on a scale representing the normal range of amplitude could aid the rapid identification of myopathies.

Amplitude also has correlations with other characteristics that may make it valuable to graph it on a two dimensional scatter plot with another characteristic. Amplitude and Duration are positively correlated. Amplitude and AAR are negatively correlated which leads to more separable data distributions when the two are plotted together (Sonoo and Stålberg, 1993).

3.3. Area

DQEMG also measures the area under the curve of the MUAP (both the positive and negative area). The area under a curve tends to be higher in neuropathy than in myopathy and lower in myopathy than in normal subjects. In the DQEMG application area is reported in the MUAP template and the Results and Summary screens, but clinical physicians do not typically use area for diagnosis. Area is displayed in units of μVmsecs in DQEMG.

3.4. Area to Amplitude Ratio (AAR)

When the area is divided by the amplitude, which is the height of the signal, the resulting number is a measure of "thickness". The AAR ranges from 0.2 to 3.5 ms. It has a fairly Normal distribution among the myopathic population, but again the range of this distribution of 0.2 to 2 ms overlaps significantly with the neuropathic distribution of 0.3 to 3.5 ms (Zalewska and

Hausmanowa-Petrusewicz, 2000). That is, a large AAR can eliminate the possibility of a myopathic diagnosis, but a neuropathic muscle might present a normal or reduced AAR. Thickness can be judged qualitatively in a visual manner in that physicians can see if the EMG signal seems "thin". In most cases the thickness of a normal MUAP falls between myopathic and neuropathic MUAPs.

AAR appears to have a reasonably Gaussian distribution within an individual. In 2 neuropathic studies conducted with the DQEMG application by Dr. Brad Watson, of University Hospital in London, the MUAP AARs within each individual were found to be closely distributed around the study mean.

3.5. Size Index

Combining the AAR with the amplitude provides an index of the MUAP's overall size. This is called the size index. It is reported to be calculated with this formula:

Size index = 2 log (Amplitude) + Area/Amplitude[1]

The size index may be a very useful MUAP characteristic for a number of reasons. Firstly, it directly reflects the physiology of the muscle. The size index has been found to depend on "both the number of muscle fibers in the motor unit territory and the size of the territory, as well as the diameter of the muscle fibres." (Okajima, et al, 1999). The separation between myopathic and neuropathic distributions is better for size index than for other measures of MUAP size, such as amplitude and duration (Zalewska and Hausmanowa-Petrusewicz, 2000). Furthermore, size index is not sensitive, as amplitude and duration are, to how close the needle electrode is to the motor unit whose AP is being detected. The size index "does not change in a numerical sense, irrespective of the distance between the recording electrode and current source (Okajima et al.1999)".

The mean myopathic size index is approximately –0.04 in biceps and oscillates around 0 due to a negative correlation between amplitude and AAR in myopathic cases. There is a positive correlation of AAR and amplitude in neurogenic cases so the neuropathic size index increases with an increase in amplitude and has a mean of approximately 2 with a much higher standard deviation

[1] Calculating size index with this formula may not produce values within the ranges described in the literature. Using μVs for amplitude, a scaling factor of log(1000) is necessary to convert units. The correct formula is:

Size index=2 (log (amplitude)-log (1000)) + area/amplitude

This scaling factor is not mentioned in Sonoo and Stålberg, (1993) but it appears in an example in Zalewska and Hausmanowa-Petrusewicz, (1999) and the reference values between them agree.

than in myopathic cases (Zalewska and Hausmanowa-Petrusewicz, 1999). The two distributions are shown in Figure 3-2.

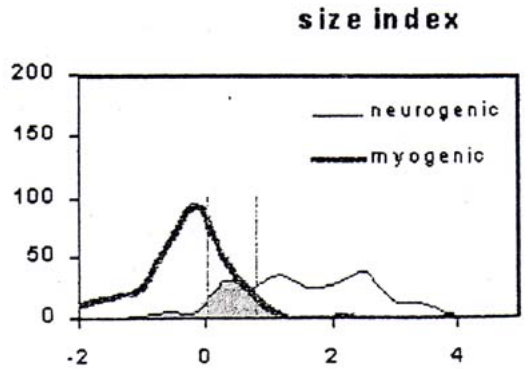


Figure 3-2 Size Index distribution in myopathy and neuropathy (Zalewska and Hausmanowa-Petrusewicz, 1999).

Since the distributions of size index in disease classes are so separable, displaying size index in a distribution histogram could be quite effective for distinguishing between abnormal states. It should be noted however that since the normal range for size index is in the middle and overlaps considerably with the myopathic and neuropathic ranges, the user should not be encouraged to use only size index to distinguish an abnormal condition.

3.6. Firing Rate

Each time a muscle contracts, a certain number of motor units are recruited to maintain the contraction. There are three variables that control the strength of a contraction: how many motor units are recruited, what types of motor units are recruited, and how quickly the motor units repeatedly fire. The number of times a motor unit fires per second is called the firing rate. It is measured in Hz.

As with all the other characteristics, firing rates vary by muscle. They also vary with the level of contraction or %MVC. Stashuk and Doherty report a normal firing rate range of 9.3 – 13.8 Hz in a distribution with a mean at 11.4 ± 1.4 for the First Dorsal Interoseus, 10.7 – 15.1 Hz (12.3 ± 1.3) in the Biceps in a low to moderate contraction. The average firing rate tends to be lower than normal in myopathic muscle studies in our lab and higher than normal in neuropathic studies. Normal Firing rates generally tend to be around 10 Hz, so if firing rates are displayed in a distribution histogram with a range from 0 to 20 Hz that would allow the user to correctly conclude that firing rates dramatically to the right or the left of middle are not normal, assuming a

moderate contraction level. In order to correctly interpret firing rate information it is necessary to know the %MVC and relate that to the firing rate. This necessity confounds our ability to assert more about firing rates; it can be noted that a low contraction level produces ample MUAPs in a myopathy but a high contraction level is necessary to sample enough MUs in a neuropathy. Thus the observed firing rate patterns in existing studies may have more to do with contraction level than the condition of the muscle.

3.7. Firing Rate per Motor Unit (FR/MU)

Taking the average firing rate during a contraction and dividing it by an indication of the number of motor units active during that contraction, one can calculate a Firing Rate per Motor Unit (FR/MU). FR/MU is a measure of motor unit recruitment. The ratio of average firing rate to the number of active MUAPs should be approximately 5 according to common belief and some literature because "by the time the first MUAP frequency reaches 10 Hz, a second MUAP should begin to fire" (p 198, Preston and Shapiro 1998). In a clinical study of the biceps, however, the normal FR/MU was found to be 2.6 on average (see the section on simulating data in chapter 7). The normal firing rate for a single MU can be between 5 and 50 Hz. A ratio of 30 Hz to 1 MUAP can indicate normal firing rate but *reduced* recruitment, thereby indicating axonal loss or conductive block (Preston, Shapiro). In other words, a neuropathy may be indicated by a high FR/MU. During a contraction in a myopathic muscle, more motor units will be recruited than normal for a given level of contraction, so the FR/MU is lower than normal.

In a myopathic case you also see *early* recruitment, where a motor unit that was formerly recruited late in a strong contraction is recruited earlier at a lower level of contraction. Many motor units may be recruited to maintain a low level of force. In order to assess early recruitment, you need to know the level of contraction, or how much force is being generated (Preston, Shapiro 1998). Usually only the electromyographer knows the level of force; the EMG practitioner typically provides resistance as the patient contracts against the elecromyographer's hand, so he can estimate the level of force and aim to keep that force consistent between contractions. DQEMG calculates the percent of Maximal Voluntary Contraction (%MVC) using the root mean square (RMS), providing a quantitative estimate relative to the level of contraction.

Actual quantification of early recruitment would be difficult, though the recruitment time, or first firing, of each motor unit could be identified. When DQEMG is reporting on firing rate, there are quite likely fewer MUAPs for which it could calculate firing rate than there were MUs active in the contraction. There are default thresholds such as amplitude and number of firing times

below which MUAPs and their trains will not be identified. Motor units that were too far away from the needle electrode or which did not fire often enough will be excluded from the study. This may dramatically alter the estimate of active and detected MUs; how much depends on the estimation algorithm. Firing rate/MUAP ratio is very hard to be confident of for this reason. Still, it is a consistent relative measure and therefore usable despite the fact that it doesn't exactly quantify recruitment.

3.8. Phases

The number of phases in a MUAP is defined as the number of times the MUAP waveform crosses the baseline, plus one, or the number of times it departs from the baseline and returns. In order to be counted, a phase also has to achieve a minimum amplitude of 20 μV and a minimum duration of 240 μs. The phase count is an indication of complexity. Generally, a normal MUAP has about 3 phases and a MUAP with more than 4 phases is considered polyphasic. Fewer than 3 phases detected in a MUAP could be an artifact of the distance of the motor unit from the electrode and is not considered abnormal.

A more complex MUAP could be caused by a number of physiological conditions. In myopathy the phase count tends to be elevated and widely variable between MUAPs. The high variation is partly due to a combination of low amplitude signals from small muscle fibers and irregular distances from muscle fibers in a motor unit to the electrode due to muscle fiber death and subsequent holes in the otherwise random distribution of muscle fibers across the motor unit territory. During axonal loss and reinnervation, there will be variations in firing while the motor endplate connection between the new axon and the muscle fibers is still being established.

Information from the number of phases is not considered to be specific in that it does not indicate a particular pathology. However, it is a sensitive indicator; polyphasic activity in 10% or more of the MUAPs from one muscle is considered a strong indication of abnormality (in all muscles but the deltoid, which has a higher threshold of abnormality). The physician usually makes a qualitative assessment of polyphasic activity, however, and does not normally calculate the actual percentage of polyphasics.

The reference values for normal phases vary according to the muscle being examined. Doherty and Stashuk (2001) report consistently lower means for phases (in the range of 2.5 to 2.8 in different muscles) than Bischoff et al (1994) reported some years earlier (ranging from 2.62 to 3.16 in the same muscles). This was particularly dramatic for the first dorsal interoseus (FDI). Doherty and Stashuk report FDI to be distributed around 2.6 ± 0.1 while Bischoff et al's value was

3.13 ± 0.38. It may be that Doherty and Stashuk define higher thresholds of duration and amplitude below which a baseline crossing is considered attributable to noise, and is not counted as a phase change.

3.9. % Polyphasic

Since the DQEMG application can count the phases in each MUAP template, it can report the number of MUAP templates that are polyphasic (phases>4) as a percent of the group of calculated templates. As with FR/MU, the precision of this percentage is compromised by the fact that some MUs will likely be excluded from the study if the application doesn't have sufficient data to calculate a template for them. If the physician does not move the needle much between contractions, it is also possible to sample the same complex motor unit multiple times, artificially elevating the % polyphasic MUAPs. However, as with FR and FR/MU, % polyphasic MUAPs is a useful relative measure.

3.10. Turns

A turn is defined as a change of slope that is maintained for a minimum of 25 μV (Stashuk, 1999). The number of turns in a MUAP is another measure of complexity. Though turns and phases may not help in differentiating between neuropathic and myopathic MUAPs, features like turns and phases could be helpful in evaluating the intensity of a pathological process (Zalewska and Hausmanowa-Petrusewicz, 2000).

More details about exactly how the DQEMG interface works and what it calculates and reports are provided in the next chapter.

Chapter 4 Analysis of the DQEMG Interface

> "Another potentially dangerous shortcut is the expert user interface review by an HCI professional (after the interface has been designed and implemented). This person's opinion may be better than a coder's, but it is still just an opinion until confirmed by feedback from representative users."
>
> Paul Smith (pwsmith@ca.ibm.com)
> IBM, Toronto Software Lab
> March 2001

Though it is of limited use, as argued above, a usability analysis of the DQEMG interface gives us a way to anticipate problems as well as guidelines for fixing them if they are confirmed by user testing.

Since 1993 when it was first written, the DQEMG application has never been user tested, and there is no testing protocol for regular testing of functionality by developers. It has never been rigorously tested for functionality. In 1999/2000 there was a preliminary review and some bugs were fixed. Comments from that review and a more recent review are merged here as they were both done by the author. A number of specific design suggestions following from this analysis are grouped at the end of the chapter under the heading "Design Suggestions."

4.1. Running DQEMG independently

When the DQEMG program starts up without having been called by the Comperio system, most of the screen is gray. There is a menu bar at the top of the screen with a window title of "DQEMG". There is a small menu selection under that and a button toolbar just below it in which most of the buttons are also gray. The user sees the following menu options:

File Options View Help

The available buttons are **Open** and **Acquire.** At this point, the user's options are fairly obvious. All buttons other than the ones listed above are grayed out, a standard way to indicate they can't be used. The buttons stand out more than the menus since they are closer to the center of the screen, they are 3-dimensional, and the text on them is larger with a thicker line.

The **File** menu's available options are mostly redundant to the button functions, but the menu option names (**Acquire New Contraction** and **Open Study**) give more information than the button names do. Additionally, there's a **Print Setup** option and **Exit** under the **File** menu.

The **Options** menu leads to **Auto Save** and **Save Raw Data**, which both turn out to be things to turn on or off, with a check mark indicating state. By default, **Auto Save** is not selected and **Save Raw Data** is. There is also a **Change Default Data Directory**. Under the **View** menu is only a "**Show all Toolbars**" option that is selected.

If the user selects the Help menu it has a solitary entry, "About DQEMG…" as shown in the figure on the right, which leads to a pop-up window listing the copyright and version information. While users might expect to see this information under the help menu, they might also be seeking help using the program, which is not there.

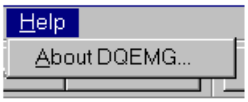


Figure 4-1 Help Menu

4.2. Opening a Study

4.2.1. Selecting a Study to Open

When the user clicks on the **Open** button or selects **Open Study** from the File menu, a pop-up window prompts him to select a muscle study to open. The DQEMG application depends on a file system where muscle study data files are stored in a folder named after the muscle, in a folder named with a patient name, in the folder of an operator. To select a muscle study the user needs the "data directory" to be the one containing the operator directory, not the one directly containing the data files. If the user needs to change the directory, he would click on the **Change Data Directory** button, which leads to a slightly confusing and non-standard interface where a small-icon directory list is in a white space framed by a gray window. A drop-down menu of drives below the file list area allows the user to change drives. The file path of the current data directory is written in black on the gray background in the lower left corner. This is a poor substitute for the input box that in more standard interfaces both allows the user to type in a location and displays the current directory as the mouse is used to change the selection. A **Prev Dir** button will take the user up one directory level and a **Default Data Directory** button will change the data directory setting to the default, which is hard-coded into the application. The window title says "Select Data Directory" and **OK** and **Cancel** buttons are the other options.

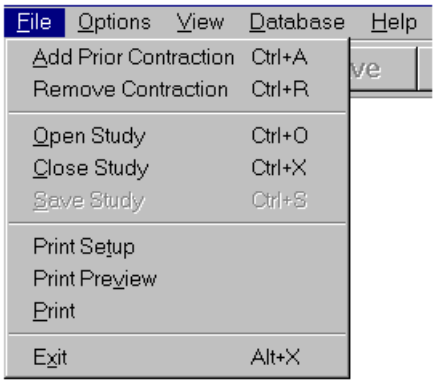
If the user clicks on the icon of a folder it will be selected. If he double-clicks on a folder, it will open **and** be selected. If he is in a folder and hits **Prev Dir** the directory containing the original folder will now be selected, but the file path in the lower left corner is the only sign of

that. There is no way to input a file path other than to browse. If the user has selected a folder at the top level within a drive, selecting the drive again with the drop-down menu will not deselect the folder. The user must click on **Prev Dir** to do that.

If the user opens a study when he already has an open study that has been modified, the program puts up a message box that asks if the user wants to save changes to the current study. This is due to the fact that it is essentially a single document (SDI) program. Response options are **Yes** and **No**. If the user mistakenly initiated a study opening (or failed to realize this was a single document program) there is no way to cancel the process at this point; even if the user cancels one step later at the window for selecting a muscle study, at the very least the current study will be closed.

4.2.2. Interface with Study Open

When the user opens a study he first sees a pop-up message, "Setting up decomp summary." There may be other messages depending on the condition of the study. The DQEMG application opens the study to the Results screen. The fact that the Results screen follows a message about a summary may contribute to summary/results term confusion discussed later. Being in the Results screen is indicated by the **Results** button, which is drawn as though depressed. All buttons along the top are available except the **Save** button which will be available once the study has been altered.



There are now five menus, and some pre-existing menus have changed. Some of the new options within pre-existing menus were there in the menu listing before the study was opened, but gray to indicate they were not available, and some options were not listed at all before the study was opened.

Figure 4-2 DQEMG menus with File selected.

The **File** menu has additionally, **Add Prior Contraction**, **Remove Contraction**, **Close Study**, **Save Study**, **Print Setup**, **Print Preview**, and **Print** options. Under **Database** is **Add Muscle Study Parameters to Database** and **Statistical Comparison**. If there is no database, nothing happens when **Add Muscle Study Parameters to Database** is selected. No feedback is given. When **Statistical Comparison** is selected the feedback is two pop-up windows, "There are no records for this studyID=0." and "Can not compare. Insufficient information available." This

area of the application is not fully developed, though it is more functional if a Comperio database exists on the computer and has an entry for the patient.

In the Results screen there are also buttons available at the bottom of the screen about the database: **Add to DB** and **Compare**. Selecting the latter of the two calls the same function as **Statistical Comparison** in the **Database** menu. Selecting **Add to DB** on the other hand has a different effect than selecting the **Add…** option under the **Database** menu. It brings up a window for Characterizing the muscle. That window has a checkbox for "Characterize muscle" and one for "MRC grade". If the user tries to continue without selecting either of these an error will inform them that they must characterize the muscle before adding/updating it to the database. They could simply be told this in the first place rather than having it presented as an option. Once both boxes are checked the scales below them become available. Though in appearance these are sliding scales, the sliders have fixed options. The Characterize Muscle scale goes from Myopathic on the left to Neuropathic on the right with Normal in the middle. The user can specify the characterization of the muscle as "severe" "moderate" or "mild" at each end of the spectrum. The MRC grade can be set to any integer from 0 to 5.

The characterization results do not seem to be saved by the DQEMG application itself. There is no way, then, for the user to record their diagnostic conclusions in this application for later review. If a user is reviewing a study done by another user, there is no way for them to look up how that other user characterized the study. The interface as it is would be extremely useful if it was always available and the results of it were made available as well.

If the characterization is done and the user tries to continue without that study having an existing patient entry in the Comperio database, the user sees the error, "There is no patient or label information for this muscle study. Saved data will be in a temp directory." However, data is not actually saved in any temp directory. Since the patient name is saved in the directory name and in study.txt, there is no reason why you should not be able to add a study to the DQEMG database even without having a Comperio database on the computer. This will be better developed in future releases of DQEMG

4.3. Acquiring a New Contraction

When the user enters the DQEMG program through the Comperio system after acquiring a sample EMG signal, the toolbar buttons and the menu options are essentially the same as when opening a study. An additional interface component exists when beginning a study, since the DQEMG application will ask if the signal just collected was a maximal voluntary contraction (MVC) and

will keep prompting the user each time a contraction is collected until an MVC is provided. The other major difference is that the program will open up into the Decomposition Summary screen instead of the Muscle Study Results screen. The **Decomp** button is drawn as though depressed. To acquire another contraction, the user would push the **Acquire** button in the top toolbar, which is nice and straightforward. Not all of the top toolbar is so clear.

4.4. Top Toolbar

In either case (opening a study or acquiring a new contraction), the top toolbar is all available once there is an open study, an exception being that the acquire button may not be available if EMG/EP is not running (i.e. there is no acquire program available), as in Figure 4-3.

Figure 4-3 Top button bar in DQEMG while on Results screen with a study open.

Each button has a mouse-over message that the user sees if the mouse indicator is paused over the button. From left to right on the toolbar, the buttons and their mouseovers (some have none) are as described in Table 4-1 below.

Open	Open Study Information
Close	Close Current Study
Save	Save Changes to Study
Results	Muscle Study Results Table
Acquire	_____
Add	Add Previous Archived Contraction
Remove	Remove this Contraction
ReDecmp	
Summary	Contraction Summary Table
Needle	View MUAP templates
Decomp	Decomp Summary
Ensemble	Muap Ensemble Screen
Macro	
EMGs	Signal Summary Screen

Table 4-1 Top toolbar button labels and mouseovers (seen in all screens).

A cursory review of the button names and mouseovers reveals that the term "Summary" is overused. It is not a very specific term. It suggests a body of information, but not what that body of information might be. Also, "Results" and "Summary" are synonyms and could be confusing to the user. The use of the word "Table" in the **Results** and **Summary** mouse-over messages does indicate that the information in them is in tabular form, but it is the terms "Study" and

"Contraction" in the mouse-over that really distinguish between these two options. Perhaps these terms should be used in the button names, which are used for navigational purposes.

If **Macro** had a mouse-over it could be "View MUAP templates" just as well as for **Needle**; the one goes to the macro EMG MUAP templates and the other goes to the needle (micro) EMG MUAP templates so the mouseovers should be more specific. The navigation information for these two options could be more complete, but at least the button names distinguish them from one another and from other options.

The toolbar buttons are organized so that things that affect or report about an entire muscle study are to the far left, functions for adding or removing a contraction in the study are in the middle, and screens for viewing and editing individual contractions are on the right. **ReDecomp** re-decomposes the signal according to the current decomposition options. If the user has not changed the decomposition options, **ReDecomp** doesn't change anything about the study except to overwrite any editing the user may have done. The toolbar organization, while logical once explained, is not very apparent to the user. There are two slight spaces in the toolbar separating the three groups of functions, but there is no label or color code to indicate the logic of the organization.

Because the DQEMG application is a single document interface, the **Open** button causes the current study to be closed. The **Close** button is therefore redundant; the only time a user would close a study without opening a new one would when finished with the application and wanting to exit. The toolbar is cramped for space; removing the **Close** button might make room for other buttons to be labeled more clearly.

4.5. Screen Designs

For the purposes of this user interface design analysis, attention is directed to the main two Screens that are used for editing and assessing the decomposition, which are the **"**Decomposition Summary" and "Needle EMG Template Data" screens, as well as to the primary information display screens, which are the "Muscle Study Results" and "Contraction Summary" screens. Since some of these screens lead to other displays or interface components that are important to the tasks they support, discussion of those subcomponents will be included as well. There is a whole section on the Marker Editing sub-screen because that area of the interface has been identified as a frustrating, problematic area by current users.

4.5.1. Muscle Study Identifiers

In all four of these main screens, a file path for the current study is displayed in the top right corner of the screen. Due to the conventions of file saving in this application, that file path reports the name of the operator (as self-defined by the operator), the name of the patient in whatever form the study operator input it to begin with, and a shortened form of the name of the muscle, which is determined in the Comperio Interface when the data is acquired. Below that line is the current day's date. There is no indication to the user of when the study was recorded from the patient, nor of when the study was last edited. The current date is not a very meaningful piece of information to have on the screen, while other information that would be useful to the user is noticeably missing, such as the date the study was acquired.

4.5.2. Decomp Screen

When the **Decomp** button is pushed, a message box comes up that says "Please wait. Drawing images to bitmap." This is a useful and friendly sounding message.

The "Decomposition Summary" or Decomp, screen displays information about the identified MUAP trains (MUAPTs) in a particular contraction. The contraction number and % MVC RMS are shown in smaller text under the title. This display format is shared between all the screens navigated to by the right-hand top toolbar, by way of the **Summary**, **Needle**, **Decomp**, **Ensemble**, **Macro** and **EMGs** buttons. These screens are also linked in the application so that the contraction in one screen will be the contraction shown in all of the other screens in this section of the interface when the user navigates between them.

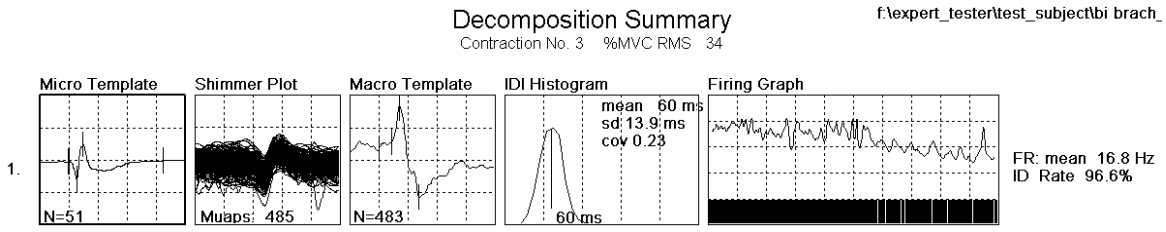


Figure 4-4 Decomposition Summary screen sample.

The view in the Decomp screen is of a row of graphs for each MUAPT, labeled at the left of the screen with that MU number. The graphs are stacked in columns, with the graph type names along the top. Left-to-right, the graphs are: Micro Template, Shimmer Plot, Macro Template, IDI Histogram, and Firing Graph. IDI stands for Inter-Discharge Interval and represents the time between a motor unit's subsequent firings. To the right of the graphs are listed a firing rate (FR) mean and an ID rate. The ID rate is the percent of MUAPS predicted by the firing pattern of that

train that were actually identified in the signal. If the ID rate is low the MU may not have been consistently active or the train may not be valid.

The main graph waveforms are drawn in yellow. Markers in the micro and macro template graphs are indicated by vertical green lines. The means on the IDI Histogram graphs are marked with a longer green line. Statistics on the graphs are in white. The graph borders are blue and the dashed gridlines are gray. Each graph is displayed about an inch and a half high.

The scales and sweeps for each column of graphs are listed at the base of the columns. If there are more than five MUAPTs in the contraction, a scrollbar appears on the right side of the view and the user must use that scrollbar in order to see the scales of the graphs.

The Decomposition Summary screen has the most complex bottom toolbar in the whole program. Because there are so many buttons for the amount of space, the buttons to change the vertical scales and horizontal sweeps of the graphs are labeled with icons.

Figure 4-5 Left half of bottom toolbar in Decomp screen: scales and sweeps.

The use of these icons for scale and sweep changes is unique to the Decomp screen in the DQEMG application. Other screens use buttons with text labels like “Scale +”. The mouseover messages for The Decomp bottom toolbar are in 4-2.

Micro	(not a button, just a label)
/\ | \/	Increase Micro Template Scale
\/ | /\	Decrease Micro Template Scale
<->	Decrease Micro Template Sweep
>-<	Increase Micro Template Sweep
Shimmer	*(not a button, just a label)*
/\ | \/	Increase Shimmer Scale
\/ | /\	Decrease Shimmer Scale
<->	Decrease Shimmer Sweep
>-<	Increase Shimmer Sweep
Macro	*(not a button, just a label)*
/\ | \/	Increase Macro Scale
\/ |	Decrease Macro Scale

/\	
<->	Decrease Macro Sweep
>-<	Increase Macro Sweep
Prev	View Prev Contraction
Next	View Next Contraction
Markers	Edit markers
Raster	Edit RasterPlot
Edit	Select and Edit a Graph
Details	View Raster

Table 4-2 Bottom toolbar labels and mouseovers in the Decomposition Summary screen. Icons approximated.

The way the "sweep" function works may be confusing if the user is not familiar with an oscilloscope, but all of the users of this program are expected to be familiar with that device and its interface. The oscilloscope was the precursor to the electromyograph and still influences its design. The sweep icons represent what will happen to the waveform if the button is pushed. Decreasing the sweep is similar to focusing in on the graph – the waveform will get wider. Increasing the sweep will shrink the information on the graph to a more narrow horizontal scale. The sweep and scale buttons are ordered so that the left-hand button will in both cases stretch the waveform while the right-hand button will shrink it. While the theory behind these icons is understandable, the scale at which they are presented makes the decrease scale icon and the decrease and increase sweep icons hard to parse. The user may tend to interpret them as obscure cryptographs rather than as composites of directional (arrow) icons. Some options are discussed in the *Design suggestions* section later in this chapter.

Figure 4-6 Right half of the bottom toolbar in the Decomp screen: Prev, Next, and mode buttons.

The buttons to the right of the scale and sweep buttons on the bottom toolbar are **Prev** and **Next**. "Prev" is an awkward abbreviation for "previous" but with the **Next** button to the right of it the meaning of the label is clear. By default after decomposition the active contraction is the last one that was collected, so the Next button will be unavailable as shown in Figure 4-3. The only way for the user to switch to another contraction is by pressing the **Prev** button. The availability of the **Prev** button is the only indication that there are other contractions available, and there is no indication on this screen as to how many contractions there are in the study. The user would repeatedly click **Prev** to find that out. The number of the current contraction is part of the screen

header, but that number comes from when the contractions were first collected; it is not an indication of place. If contractions have been removed from a study the user could find that contraction number 5 is the third contraction in the study, following contraction number 2.

When it is brought up for the first time, the decomposition Summary screen is in a non-interactive mode, so pointing and clicking within the screen does nothing. Pressing any one of the **Markers, Raster**, **Edit** and **Details** buttons will turn on an interactive mode. When the mouse is moved over a graph that can be selected in this interactive mode, the rectangular outline of the graph will be highlighted in a magenta color. There are two ways to exit an interactive mode. Either the user can push the escape button (**Esc** on the keyboard) or the user can select a different mode. Pressing escape is the only way to go back to no mode once a mode has been selected. This is not documented anywhere and the user is not expected to figure it out.

"Markers" mode allows the user to enter one of the marker editing screens for the macro or micro templates by clicking on that template graph. The template graphs are the only graphs that can be selected in this mode. Clicking on a graph sends the user to a different screen, from which the user has to select a **Close** button in the bottom right corner of the screen or the **Decomp** button in the top toolbar in order to return to the Decomposition Summary screen.

"Edit" mode is used to set a graph or MUAP train invalid, so that its statistics are not included in the contraction and muscle study statistical tables. Any graph can be selected in this mode and will turn gray to indicate it is invalid. The IDI Histograms and Firing Graphs can only be selected together since they are both based on firing pattern information. The whole train and all its graphs will be highlighted if the mouse pointer is moved to the right or left of the row. This mode does not send the user to any other screens.

"Raster" and "Details" mode can both be used to view the Shimmer graph; other graphs cannot be selected in those two modes. The Shimmer graph is an overlay of all the MUAP waveforms that have been assigned to one train. The Raster plot is just like the Shimmer graph except that the waveforms are drawn with a small vertical separation, with the earliest waveforms at the top of the screen and time proceeding in the down direction. Clicking on a Shimmer graph while in "Raster" mode lets the user edit individual waveforms out of the train. Clicking on a Shimmer graph while in "Details" mode takes you to a screen where the vertical expansion is even larger and some unexplained numbers label each waveform. Both of these modes take the user to a different screen with a type of raster plot on it. One lets you edit the MUAP waveforms, the other does not. The only way to get from the editing screen to the detailed screen is to go back through the Decomposition Summary screen, select the other mode, and reselect the shimmer graph.

When the user is in a sub-screen such as the raster plot, there are two ways for the user to return to the parent screen. The user can select the **Close** button somewhere in the bottom button bar to return to the parent screen, or he can click on the button of another major screen and then back on the button of the parent screen. It is anticipated that if the user does this rapid back-and-forth between screens in order to get to a parent screen perhaps the architectural model of "closing" a sub-screen that is not drawn as a pop-up window will not match the user's mental model of interface navigation.

Also, the **Close** button may be difficult to find even when the user anticipates using it. It is currently located at the right end of the bottom toolbar, but depending on the length of the toolbar, that could be anywhere from the left side of the screen to the right side of the screen. It would be better for the location of this button to be completely consistent between screens.

4.5.3. Needle Screen

When in the "Needle EMG Template Data" screen, the **Needle** button appears depressed on the top toolbar. The title of the screen is in the center in large white letters with the contraction number and % MVC RMS below it in smaller letters. The file path and current date are again in the upper right-hand corner. The upper left-hand corner has an ominous-looking red text notice: "MACRO Neg Peak Onset shown in RED." The main part of the screen shows the MUAP template graphs (see Figure 4-4). The default scale units are 200μV per division. Figure 4-4 (next page) shows a large MUAP at this scale; the fact that it is larger than normal is immediately apparent.

In the bottom left-hand corner of the scrollable part of the screen are the scales for the graphs:

Horz: 5 ms/div
Vert: 200 μV/div
Sweep: 25 ms

Each row of template graphs has up to three graphs in it. If there are more than six template graphs in the contraction, this part of the screen has a scroll bar on the right and the user has to scroll in order to see the scale information. It would make more sense to put the display scale key at the top of the screen.

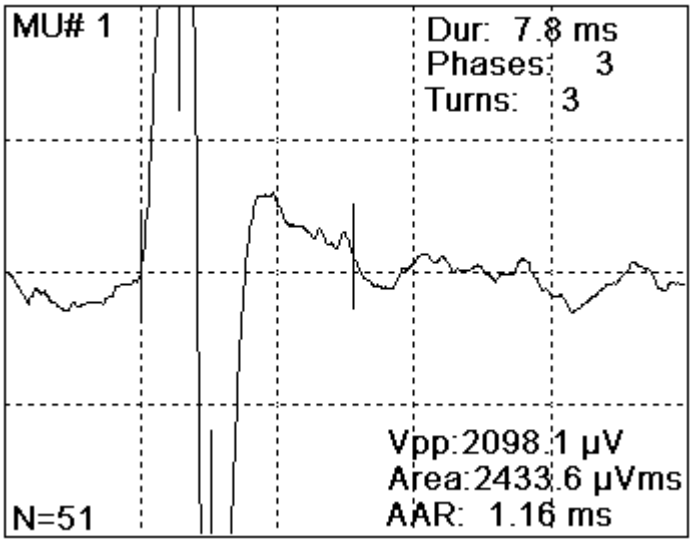


Figure 4-7 A MUAP template graph in the Needle screen.

4.5.3.1. Bottom Toolbar

The **Prev** and **Next** buttons are in the middle of the bottom of the screen. The order of the available buttons along the bottom of this screen, and their mouseovers, left-to-right are listed in Table 4-3.

Scale +	Increase Vertical Scale
Scale –	Decrease Vertical Scale
Sweep +	Increase Sweep
Sweep –	Decrease Sweep
Draw All	Show All Muaps
Valid	Show Valid Muaps
Prev	View Prev Contraction
Next	View Next Contraction
Markers	Edit markers
Edit	Select and Edit a Graph

Table 4-3 Bottom toolbar labels and mouseovers in the Needle Summary screen.

The Scale and Sweep buttons change the vertical and horizontal scaling of all the template graphs. Indication of scale is by µV per vertical division. The default scale is 200 µV/div. That goes from 1,000,000 µV/div up to 0.01 µV/div. Anything from 20,000 µV/div to 1 million µV/div is essentially showing the user a flat line in each graph. Sweep goes from 50 ms to 1 ms; 1 ms is an ineffectual sweep since the MUAP templates can't scroll; the interesting part of waveforms are mostly not displayed at this sweep, or even at 2 ms. These low-level sweep options are not in there by design. They are a side effect of the architecture of the software that

runs the application. There is a set of scale levels used by many parts of the application. However, it is unlikely that the user will be bothered by this effect. The main user-centered design concern on the sweep and scale buttons is whether the user's mental model will match the function of the buttons well enough for the user to consistently choose the correct + or – button to cause the desired effect on the template display.

By default the Needle screen is in a non-interactive mode so pointing and clicking on the MUAP templates does nothing. If the user clicks on **Edit** in the lower right-hand corner of the screen, the screen is in "Edit" interactive mode. When the mouse is moved over a template graph in "Edit" or "Markers" mode, the edge of the graph will be highlighted in magenta. Clicking on a MUAP template graph while in "Edit" mode causes it to be invalid and thus disappear because the screen starts off by default in "Valid MUAPs " display mode.

The display mode is indicated by the fact that the **Valid** button is selected in the bottom toolbar. The editing action in the Needle screen makes the MUAP template invalid but does not affect the validity of the whole MUAP train. If the MUAP template that was edited out was in the lower right-hand corner there will now be a gap where that template was. For all other MUAP templates when the user clicks on them there is still a template in that space because all templates shift left and upward so there are never any gaps in the middle of the display, just in the lower right-hand corner. This means that as templates are edited into invalidity the location of the other templates on the screen changes. If MUAPs have been edited out (either in the needle screen or previously in the Decomp screen), the user might notice that based on jumps in MU number sequence: MU #1 template may be directly to the left of MU # 4 template.

If the user edits out a template by mistake, he has to click on **Draw All** in order to see the "edited" template. When the "Draw all MUAPs " display mode is in effect, invalid MUAP templates appear in gray. If the user selects the "Edit" interactive mode while **Draw All** is selected, he can click on a gray MUAP template and it will be valid again. It is not obvious that the **Draw All** and **Valid** buttons represent a display mode toggle switch.

It is more common in most interfaces to select something and *then* issue a command on it rather than to enter a command mode and have something done every time you click on something. Normally in Windows you can click to select something anytime without anything happening, and you double-click on something to edit or open it. It's not clear whether the user will likely want to "edit" (delete) many MUAPs at one time that he did not already edit in the Decomp screen. It's also not clear if the value of clearing invalid MUAP templates off the Needle display outweighs the disadvantages of a non-intuitive interface.

If the user is in an interactive mode on the Needle screen and uses a top toolbar button to go to another screen, when they return to the needle screen they will still be in that mode. By design, the **Marker** or **Edit** button will be shown pressed in, so as to indicate the mode. While it is common in interface design for a button to indicate a mode (e.g. the bold button in Microsoft Word) it is also common to redundantly display status or mode information in a consistent location on the screen. That might be useful in this case, if the modal system is to be retained.

4.5.4. Marker Editing Screen

The "Needle Marker Editing" screen displays a single MUAP template. The MU number is shown to the right of the title. The contraction number and %MVC RMS are still below the title and the right-hand corner of the screen has the day's date and an indication of the muscle and patient under study.

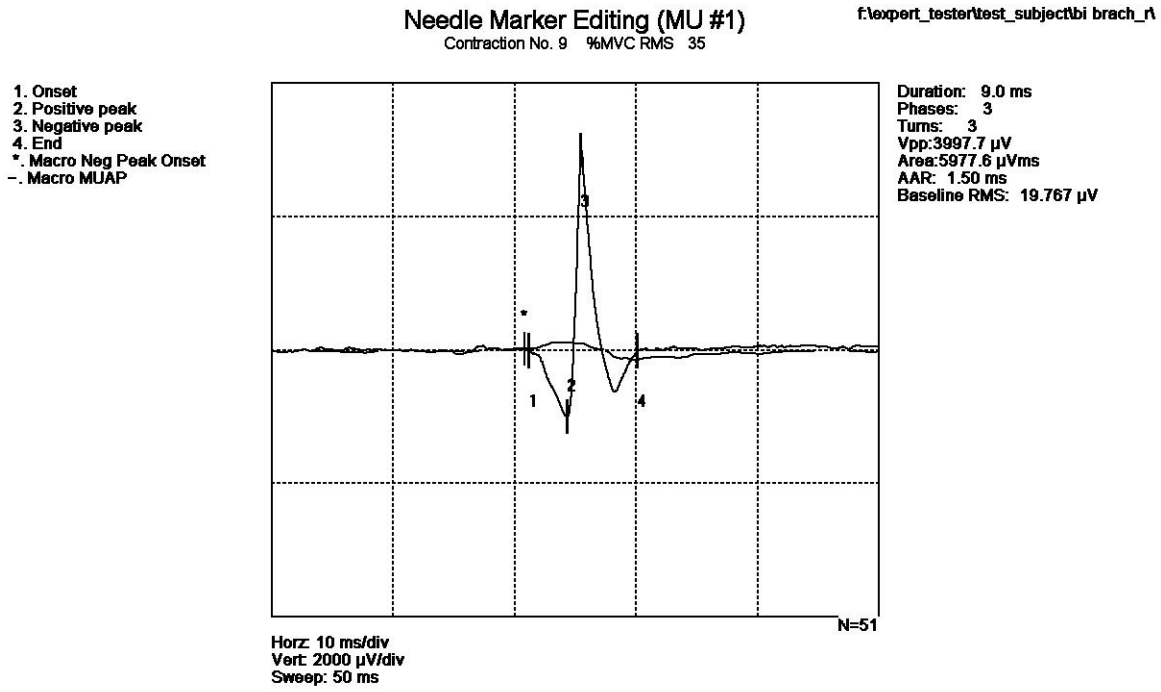


Figure 4-8 Needle Marker Editing Template

The MUAP template graph looks the same as in the previous screen except that it is larger and now the markers are labeled with the numbers 1 through 4. A key on the left side of the screen identifies them as

1. Onset
2. Positive Peak
3. Negative Peak

4. End

The key also tells the user that the red marker with an asterisk at the top marks the negative peak onset of the macro signal from this MU, and the gray line in the graph represents its macro MUAP waveform. The statistics that were superimposed on the graph in the Needle screen are now listed to the right of the graph.

Marker editing is done with the mouse. The user can click directly on the marker to be edited or on one of the marker buttons at the bottom of the screen.

Scale +	Increase Vertical Scale
Scale -	Decrease Vertical Scale
Sweep +	Increase Sweep
Sweep -	Decrease Sweep
Onset	Edit Onset Marker
Pos Peak	Edit Positive Peak Marker
Neg Peak	Edit Negative Peak Marker
End	Edit End Marker
Macro	
Prev	Previous Page
Next	Next Page
First	First Page (Home)
Last	Last Page (End)
Close	Return to Previous Screen

Table 4-4 Bottom toolbar labels and mouseovers in the Needle Marker Editing screen.

If part of the yellow MUAP waveform that contains a marker is off the graph due to the scale there is a red message just above the graph: Warning: one or more markers are off the graph. If a marker is off the graph its button on the bottom toolbar will be gray and the user can not edit it. Changing the scale or sweep of the graph will usually bring all the markers in view.

When the user points the mouse at a marker it will turn white to indicate the mouse is in range to select it. If the user clicks on it then, the marker returns to green and a red arrow appears above the marker, indicating it is selected. A magenta shadow representation of the marker appears where the marker was located before it was selected and will stay there with the red arrow until the user clicks the mouse on the graph again. During this time, the mouse function is completely taken by the template editing graph. The mouse arrow is gone and the user cannot move the mouse outside the graph. Keyboard functions such as Alt-F to activate the file menu do not work although Alt-tab still works to switch to a different program. Other than Alt-tab, the user has no way to exit the template graph editing function other than by clicking on the graph. If the user wants to change the sweep or scale of the graph, the marker must be clicked back in place first and then reselected after the scale change.

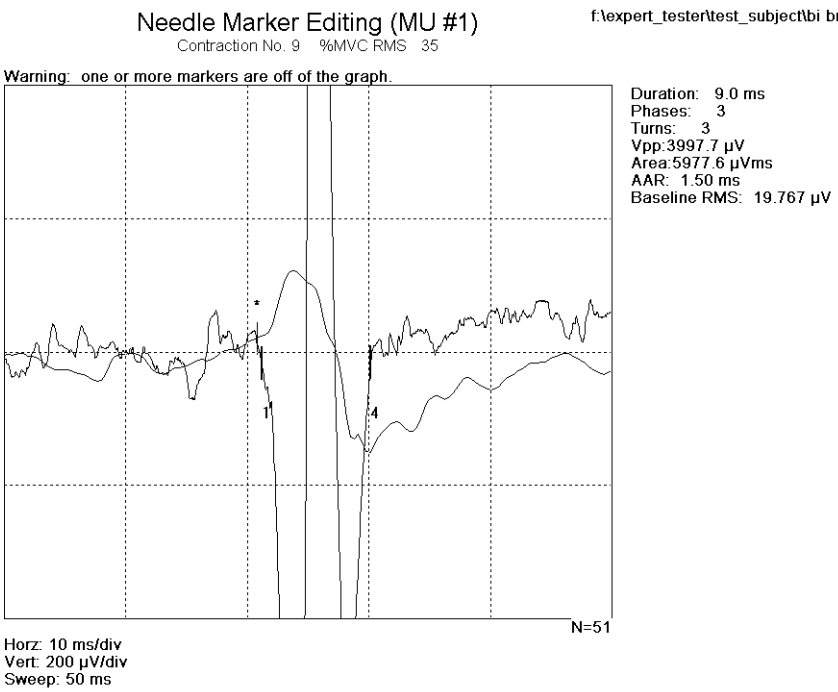


Figure 4-9 Needle marker editing template shown at default scale of 200 µV/div.

Once a marker is selected, it will move when the mouse is moved. The vertical location of the mouse does not matter in moving the marker; though the marker is moving along a two-dimensional line, the user is really only controlling the horizontal position of the marker. A vertical mouse movement does not affect the marker placement. This is non-intuitive since the mouse arrow was replaced by the marker with a red arrow over it. The user might naturally expect to be clicking and dragging this item or to have the freedom to move it in all directions such as one can normally do with the mouse.

Because of the limited horizontal movement, the tester tried moving the marker with the arrow keys while it was selected. This was not intended by the programmers and does not work. The up and sideways arrow keys do nothing but the down arrow key switches the screen to the next MU in the contraction. Similarly, the Page Up key switches the user to the previous MU but the Page Down key does nothing. When the screen is switched with the Page Up or down arrow key, it is still in marker edit mode on the marker number that was selected for the last MU. The *design* suggestions area later in this chapter explains some suggestions as to how to make purposeful use of these keys.

This screen has the same Scale and Sweep button design (and issues) as the Needle Summary screen. If you hit scale + the scale goes down while the waveform gets bigger. The **Previous** button turns gray when the user is viewing the first available MU template in the

contraction, and the **Next** button turns gray when the user is on the last template. This is good feedback. The **First** and **Last** buttons, however, are never turned gray or made unavailable, so it is possible to be on a MU with **Next** unavailable but **Last** available. In this case if the user clicks on the **Last** button, the waveform and markers flicker once but there is no other change.

There is no way to proceed to editing the markers of the next contraction other than to hit **Close** to close the Markers Editing screen and go back to the parent (Needle or Decomp) screen and hit **Next** or **Previous** there to switch to another contraction. This makes editing all the Markers in a study a tedious process. The user can also select a screen button from the top toolbar, but the button for the parent screen will be grey and depressed so it cannot be selected until the user has first gone to another screen.

Note that according to the convention of the needle EMG profession, a "positive" voltage difference peak is shown deflecting in the downward direction, while a "negative" peak deflects up above the baseline. This is not expected to be confusing to an EMG practitioner.

4.5.5. Contraction Summary Screen

The "Contraction No.__ Summary" screen presents a table of mean and standard deviation statistics for a single contraction. If there are multiple contractions in the study, clicking on Summary displays the summary of the active contraction and the user can view the summary of other contractions by using the **Prev** or **Next** buttons in the bottom toolbar. Because the bottom toolbar in this screen is not long, those buttons are in the bottom left-hand corner of the screen. It is uncommon within the DQEMG application for such navigational buttons to be on the left, but it's not clear whether or not the user will have trouble finding them. The existence of **Prev** and **Next** buttons is consistent throughout the program, so the user may expect them enough to find them wherever they are.

The Summary and Results screen formats share code, which one can tell when looking at a summary screen. Each summary is only going to be displaying results from 1 contraction. What it says at the bottom of the screen is "Results are from 1 contractions containing __ valid motor units." The 's' on 'contractions' should be hidden if there is only one. The 's' on 'units' is fine; the screen needs at least 2 valid motor units to be constructed. Otherwise the user sees a black screen and just a red error message explaining that this is the case.

4.5.6. Muscle Study Results Screen

The "Muscle Study Results" screen is a table of means and standard deviations like the summary screen; it summarizes the results of the entire muscle study instead of just one contraction. The bottom of the display indicates the number of contractions and valid motor units represented. An example of a normal patient's Results screen is included in Figure 4-10.

The information displayed on the results screen is in columns of data, organized into groups of characteristics. The group heading of Micro, Macro, IDI, FR, or Misc. is on the left of the display. For each characteristic in each group, DQEMG reports the muscle study mean, the standard deviation, and the number of samples.

Muscle Study Results \1st dor inter_r\

Group	Parameter Type	Mean	Std. Dev	# of Samples
Contraction	Percent MVC RMS	6.28 %	1.60	4
Micro	Peak to peak voltage	743.46 µV	479.59	26
	Duration	10.26 ms	4.65	26
	Phases	2.27	0.96	26
	Turns	3.00	1.33	26
	Area to Amplitude Ratio	1.33 ms	0.36	26
Macro	Peak to peak voltage	137.67 µV	123.23	27
	Area	377.41 µVms	270.85	27
	Neg. peak amplitude	82.13 µV	69.15	27
	Neg. peak area	170.88 µVms	127.79	27
	Neg. peak duration	19.13 ms	6.41	27
IDI	Mean IDI	96.44 ms	13.80	22
	IDI Standard deviation	11.43 ms	3.05	22
	IDI Coefficient of Variation	0.12 ms	0.00	22
FR	Mean firing rate	10.55 Hz	1.34	22
	FR Mean consecutive difference	0.12 Hz	0.00	22
Misc.	Identification rate	72.60 %	22.15	22
	Number of muaps	173.04	101.10	27

Results are from 4 contractions containing 27 valid motor units.

Figure 4-10 Muscle Study Results screen in DQEMG.

Number of samples is a general term, since it could represent a number of contractions, MUAP templates, MUAP trains, or motor units. In Figure 4-10, for example, the mean percent MVC RMS is calculated from a sample of four contractions, while the Micro statistics are calculated from 26 valid MUAP templates. It could be confusing to the user that the number of samples for Micro, IDI, FR and Misc. statistics is lower than the number of valid motor units reported at the bottom. There is a lot of wasted horizontal space in this display; there is room to add units to the number of samples.

The Results screen has a bottom toolbar that differentiates it from the Summary screen. There are no **Prev** or **Next** buttons, since the display is summarizing the whole muscle study.

There are buttons for adding, updating, or comparing the study to a database, the functions of which were discussed above in the *Interface with Study Open* section of this chapter. In later versions of DQEMG there is also a button that leads to a sub screen for doing Motor Unit Number estimation (MUNE).

4.6. Design Suggestions

4.6.1. Overall interface

These design suggestions follow from the interface analysis and were noted before the user testing. Design suggestions from the user testing are in Chapter 9.

4.6.1.1. Menus

- All menus and all options within them should be listed in all states of the program (study open or not, for instance). Unavailable options should be faded out in gray.

4.6.1.2. Opening a Study

- When "open study" is selected and the program already has a study open that has been modified, the options for the "save changes…?" prompt should include **Cancel** for the person who hit the **Open** button accidentally or did not realize this was a single document program.
- The file location display at the bottom left corner of the change data directory window should be replaced with an input box at the top left of the window where the file path is displayed and can be edited directly.

4.6.1.3. Top Toolbar

- The term "MUAP" should be consistently in all caps as that is the industry standard.
- Given the possible parallelism of the **Needle** and **Macro** screens, and the fact that the templates shown in each are called "Micro Templates" and "Macro Templates" in the Decomp view, the **Needle** button should be renamed **Micro -**or- the micro templates should be consistently called Needle Templates.
- The **Close** button should be removed to eliminate confusion and free up toolbar space.
- Given the possible confusion between the synonymous **Summary** and **Results** buttons, the first should be renamed **Contraction Stats** and the latter should be renamed **Study Results**. Alternatively, this toolbar should have two labels. The first, "Study", applies to

the first 7 buttons in the bar. The second label, "Contraction", applies to the other buttons. These labels could be formatted like the Micro Shimmer and Macro labels in the Decomp bottom toolbar.

- When the **ReDecomp** button is pressed, the user should be prompted to change the decomposition options. There should also be a warning that re-decomposition will remove any editing that has been done.

4.6.1.4. General Comments

- The label "Edit" should be replaced with the term "Exclude" on all such buttons on the interface and the mouse-over should be changed to "Exclude a Graph from Study Results". In accordance with this, the instructions on the Raster screen should be "Exclude selected MUAPs". Alternatively, "Edit" could be changed to "Remove."

4.6.1.5. Decomp Screen

- After a decomposition or re-decomposition, the default contraction to come up in this screen should be the first one, not the last one.
- The bottom toolbar sweep and scale icons should be replaced. Icons could be designed to incorporate the + and – signs that are used on other screens; example waveforms could be shown in the icons as well, so the scale plus button shows a short waveform and the scale minus button shows a tall waveform and so on with the sweep buttons. Even switching to using a single up arrow for increase scale, a single down arrow for decrease scale,
- The **Markers** mouse-over should say "Edit Markers on a Template"

4.6.1.6. Needle screen

- Use a different highlight color or integrate an icon into the highlighting process in order to make the Edit/Markers mode toggle more visible. If the Edit mode highlight included a big X across the graph, for example, it would be obvious the graph was being eliminated if selected.
- The use of the word "Edit" is ambiguous here. See *General Comments* above regarding changing the label to Exclude.
- The default display mode for this screen should be Draw All. This would help prevent confusion if MUAP templates were accidentally edited out, and it would be consistent with the Decomp screen.

4.6.1.7. Marker Editing screen

- An undo or reset function may be useful here, so the user who makes an editing error can undo the last action or return to the original settings.
- This screen should be scrollable across the MUAP template duration.
- The **Prev** and **Next** button mouseovers should be "Previous MUAP" and "Next MUAP".
- The **Close** button mouseover should be "Return to the Needle Screen" when the needle screen is the parent, and 'Return to Decomp Screen" when that is the parent screen.
- There ought to be a way to move on to the next contraction in the study while editing MUAP templates without having to return to the Needle screen. This could be a button for next contraction that only appears when the last MUAP in the current contraction is on the screen.
- A jump bar such as the one below might be more useful than **Next**, **Prev**, **First** and **Last**, or perhaps in addition to them. Each number would bring up that motor unit number when selected (the numbers could be formatted like buttons). This design has the advantage of giving the user an overview of how many motor units are in the contraction.

  MUAP # 1 2 3 4 5 6 7 8

  These days a control button set that makes use of VCR button conventions, such as

  << < > >>

  First Prev Next Last

  is very common in interface design. This would be a nicely compact replacement for **First**, **Prev**, **Next** and **Last**, that would get the idea across graphically.
- There should be a way to use the keyboard to select a marker to edit. Perhaps *Alt-(#)*.
- When in marker editing mode, the arrow keys should be enabled to move the marker to the left or the right.
- It may also be convenient to use keyboard shortcuts to change the sweep and scale. This keyboard function should still work even while a marker is selected and being moved.
- While the up and down arrows could be used to move between MUs, the user should not already be editing a marker after making this switch.
- Since sweep can't go higher than 50 ms or lower than 1 ms, Sweep + and Sweep – buttons should gray out at those settings.

4.6.1.8. Summary

- The Summary screen, and possibly the Needle and Decomp screens also, should indicate more specifically how many contractions there are. This could be done as suggested for

MUAP numbers in the marker editing, or it could be done in the screen header. The title of the screen could say MU #3 (of 4).

4.6.1.9. Results

- Units should be added to the number of samples to reduce confusion.
- The date the study was first collected should be displayed.
- The user should be able to input and access notes and information about how the study would be characterized.

4.6.2. New Information Display Designs

4.6.2.1. Distribution Histograms

The original Muscle Study Results display in the DQEMG application was a table of means and standard deviations. Each row also listed how many MUAPs or MUAPTs (N) were involved in the calculation of that row's statistics. Examination of distributions of MUAP characteristics in real cases, however, did not support the assumption that each characteristic measure was part of a Normal, or Gaussian distribution around a representative mean. Also, drastically different patterns of data could result in the same (or a similar mean) and standard deviation. The value of the reported statistics could be clarified by a graphical representation of the data itself (Tufte, 1983). Furthermore, the presence of outliers may in itself be a significant indicator of disease (Stålberg, Bischoff, and Falck 1994). Outliers are difficult to observe given the mean and standard deviation but they are directly visible in a distribution graph.

When the number of data points is large enough to graph but too small for calculating the distribution function, the distribution function can be approximated, or estimated, by a histogram. The type of graph used for this histogram display is a frequency polygon. A frequency polygon is constructed the same way as a frequency histogram except that points are plotted at the midpoint of each bin, at a height proportional to the frequency of that bin, and then connected together by straight lines. This is in contrast to the bars plotted at respective frequency heights in a traditional histogram. It is better to use a frequency polygon than a histogram when comparing the shapes of two or more frequency distributions (Croft 1976). Since the underlying diagnostic task consists of comparing distributions to normal distributions, and the additional disease monitoring task consists of comparing distributions over time, this is considered to be a desirable advantage. It is customary to smooth frequency distributions of large numbers into distribution curves, but in this case the number of samples is not expected to be high. There is one sample collected for each

MUAPT in the study, and that number tends to be somewhere just above 20 and probably lower than 50. Another reason for using a frequency polygon for new histograms in the DQEMG interface is that it is consistent with the frequency polygon already in use for the IDI histograms in the decomposition screen.

To make the details of the distribution more directly visible, the actual data points were plotted on the horizontal axis. The mean of each histogram was indicated with a vertical line, labeled at the top with the numerical mean. The number of samples contributing to the graph, N, is shown to the left of each histogram.

To give context to the information in the distribution, ranges for each characteristic were chosen so that the middle of the graph would represent a normal range of values and data to the extreme edges of the graph could be correctly identified as abnormal. On a practical note, we had to decide between plotting data points and distributions off the graph if they went beyond this range, and changing the range if necessary when a particularly large data point demanded it. This was particularly important for amplitude; while a normal amplitude mean might be between 300 and 500 μV, individual data points could be significantly larger than 1000 μV, even in normal cases. In neuropathic cases they could get above 3000 μV. For the prototype display in the DQEMG application, this issue was addressed by making the default range the preferred one while expanding the range by a round whole number if the range was not yet large enough to display all the data.

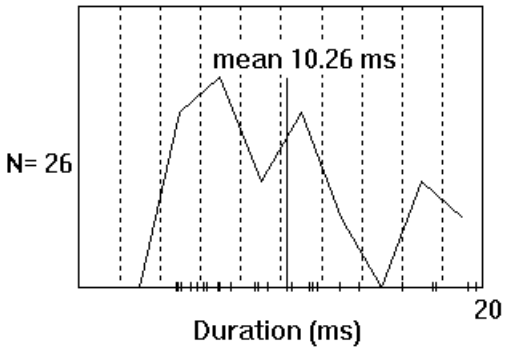


Figure 4-11 Prototype histogram for duration (normal distribution).

Colors

In order to be consistent with the rest of the DQEMG application, the axis, gridline and label colors were set to blue, gray and white, respectively. A study of color was not a major focus of this project. Data points were marked in a yellow color of a slightly brighter intensity than the blue axis, and the function line of the histogram approximation was also in this yellow color. That is the color used for waveforms and functions in the rest of the program. The mean of each

distribution was marked by a long vertical line in green, chosen because it was a visible light color that had previously been used for the markers in MUAP templates and therefore had an established function marking features of the waveform. In this case, the green line marked a feature of the distribution function rather than a landmark on a MUAP waveform, but the purpose is a parallel one.

4.6.2.2. 2-Dimensional Scatter Plots

When a relationship between two things is what is interesting to the user, a display should calculate that relationship and display it directly (Wickens, 1996). Therefore, if the covariance between two characteristics, such as the Amplitude and the Area to Amplitude Ratio (AAR), gives the user a specific way to distinguish between myopathic MUAPs and neuropathic MUAPs, as suggested by Sonoo and Stålberg (1993), then that information should be expressed to the user. The scatter plot of two MUAP characteristics can make it possible for the user to infer this relationship, and we can also relieve that cognitive task by calculating and displaying the correlation coefficient directly.

Since a 2-dimensional scatter plot of amplitude and thickness (AAR) is supposed to follow a separable pattern in neuropathy and myopathy, that scatter plot was included in a prototype graphical display, with amplitude up the left-hand axes and thickness along the horizontal axis. This orientation should correlate with the user's mental model for these two characteristics, with amplitude as something that rises and falls in strength and thickness as a sideways measure.

4.6.2.3. Polar Star Plots

A polar star display is a type of ecological interface display that is used when one of the design goals is for the state of the system represented by the display to be readily apparent to the viewer (Trujillo and Schutte, 1999). A polar star display is designed so that different properties of the system are plotted along a set of axes with the scales normalized so that a "normal" condition for the system results in the indicators being equidistant along the axes. Each variable is plotted on a radial axis from the common origin and with the range of the axis chosen to be twice normal. When lines are drawn between the axes so that a line from each property connects to the indicator for each neighboring property this results in a regular polygon, in a normal state. An abnormal state produces an irregular polygon. This is like the vertical line of the clocked dials example in Figure 2-7, only this time the shape of the polygon in the display is the emergent feature. Geometrical shapes are something people can recognize very quickly and without much load on working memory (Gibson, 1986).

The normalized axes of the polar star display provide context for the viewer. Even without the other axes, the viewer can readily identify whether a system property is above or below a normal value, by comparing the position of the displayed value to the midpoint of the axis.

A Polar Star Display for DQEMG

In applying the polar star design to the DQEMG interface, we had to select which variables to display and how to display them. The main information system states we'd like to make visible are normal, neuropathic, and myopathic. An ideal program would have statistics on the ranges of each of these states for each muscle that might be studied, possibly anticipating variation within that muscle according to age and intensity of contraction. Unfortunately, this lab does not yet have that information. Still, base on reported literature values, we constructed a general model for these ranges (and examined reported ranges in a single muscle, the biceps brachii) and then discussed how a polar star plot ought to be designed for each characteristic to be included.

There are other issues beyond the ranges of properties to be considered. Which properties of the DQEMG information set should be displayed in a polar star design? There are a number of properties that are typically used to characterize a MUAP. Duration, amplitude, size index, and area/amplitude ratio are all used to characterize the size of the MUAP. Phases and turns are used to identify MUAP complexity. The size characteristics are considered a more specific type of characteristic than complexity, since both neuropathic and myopathic muscles are likely to produce complex MUAPs. Firing rate or FR/MU could also be used, to characterize the level of motor unit activity and recruitment.

Translated into polar star design, the locations of the indicators for both phases and turns would stretch the polygon larger than normal during whichever type of abnormal condition might be going on. Including both characteristics would be redundant to the shape of the polygon, so we decided to use only one of them. There is at least one study that shows radically increased turns to be more indicative of neuropathies than an increased number of phases (Stewart et al. (1989, as cited by Zalewska and Hausmanowa-Petrusewicz (2000)), so turns was chosen over phases. Since the number of turns is a count, it does not seem too significant what angle it is oriented at, but it may fit the user's mental model best if a higher number of turns raises the indicator in the true vertical direction on the display (Wickens, 1996).

Of the size characteristics, duration and amplitude are most commonly used for diagnostic purposes. Any EMG physician will be accustomed to interpreting these characteristics, so they likely ought to be included. Amplitude, which is a measure of voltage difference, again measures something that is usually indicated along a vertical scale. Duration measures time, which is often plotted along a horizontal axis with higher values to the right. Following these conventions,

indicators for duration and amplitude were put along the right-hand horizontal axis and the nearest axis counter-clockwise from that, respectively. Depending on how many axes there are in the end product, the amplitude axis will be at some positive angle from duration.

If the polar star had just three axes, there would not be room to display much variation in shape for the polygon, which would then be a triangle. Also, the two characteristics that were already chosen for the right half of the polar star generally have higher values in neuropathy and lower values in myopathy. If another property that increased for neuropathy and decreased for myopathy was added to that side of the graph, there would be a dramatic inflation or deflation that would imbalance the polygon. This kind of shape would be readily recognizable.

The firing rate is a property of the entire MUAP train, and the DQEMG program is fairly unique in being able to present statistics on this to the viewer. Neuropathic patients will present with higher firing rates and fewer active MUs at a level of contraction where myopathic patients will show an increased number of active MUs operating at normal to increased firing rates. At a future time when the program's capacity for estimating the number of active motor units has been improved, it might be able to present some more useful analysis of recruitment. For the moment firing rate was included on the lower right side of the graph.

To complete the goal of producing unique polygons for the two major disease states, we aimed to create an asymmetrical display in neuropathy, compared to a mostly reduced myopathic display that may show a high point only along the turns axis.

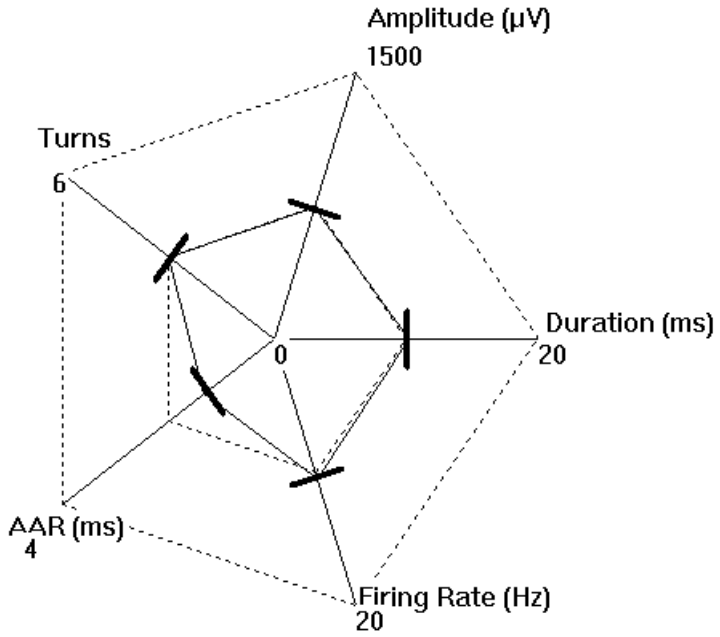


Figure 4-12 The 5-dimensional polar plot for a normal subject.

To strengthen this design, another characteristic that is smaller in myopathies than in neuropathies was added to the left side of the graph: the area to amplitude ratio, or thickness. The size index might have been even more clearly specific here. That was explored in the Polar Star plots for the Nonexpert testing (see Chapter 8) but was not included in the prototype that was presented to the Expert Testing participants, because it was assumed they would be more familiar with the characteristic patterns of AAR than with those of size index.

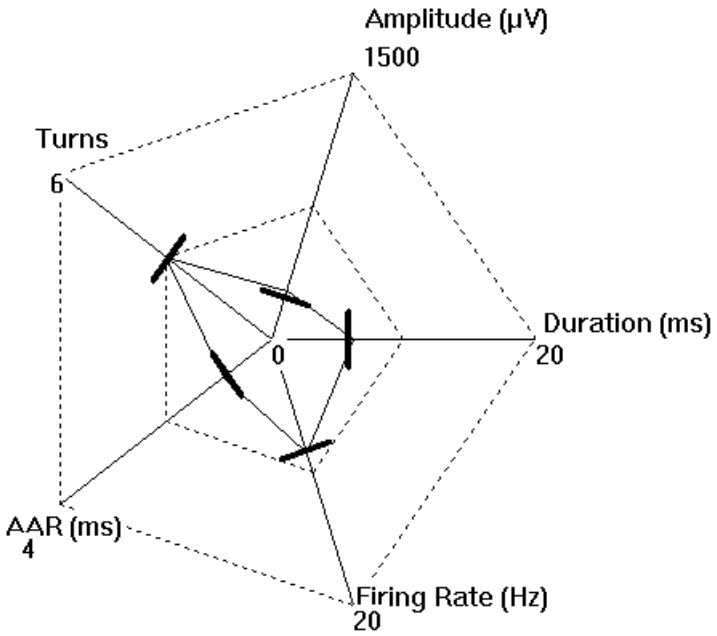


Figure 4-13 The 5-dimensional polar plot for a myopathic subject

One other pattern considered for the polar plot involved this same design but with turns in the place of firing rate. It was hoped that the emergent shape would then be more like an ellipse or diamond that would be vertically long for a myopathic patient and merely large for a neuropathic case. However, durations in myopathic patients are also commonly "normal" so the myopathic star would not consistently collapse on the right side. Similarly, the number of turns can average to a normal level in either myopathy or low-level neuropathy, so it was thought that more information was necessary to provide a specific emergent shape.

Once the set of characteristics was determined and ranges to display those characteristics were chosen to provide the properly normalized axes, it was necessary to decide how to represent each characteristic. As has been discussed regarding the other two types of graph, the mean of each characteristic may not be a good representation of how that characteristic is distributed within a particular individual. In some cases the distribution itself is more clearly indicative than the mean or any other statistic. However, with the polar plot there will be so much information on

one graph that the designer must be careful not to make it too crowded. A graph should aim to be concise, just as the written word (Tufte, 1983). It was determined that the prototype would show just the mean value for each parameter. A thick line crosses the axis at the mean, as shown in Figure 4-14.

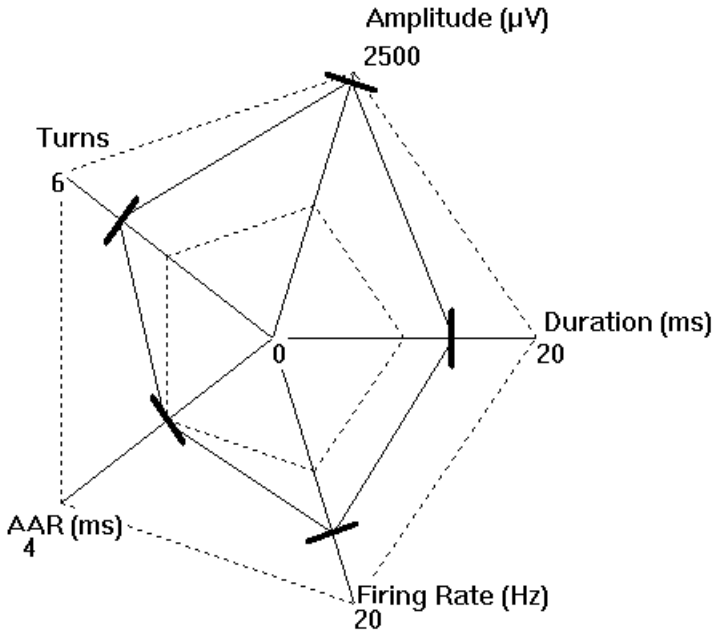


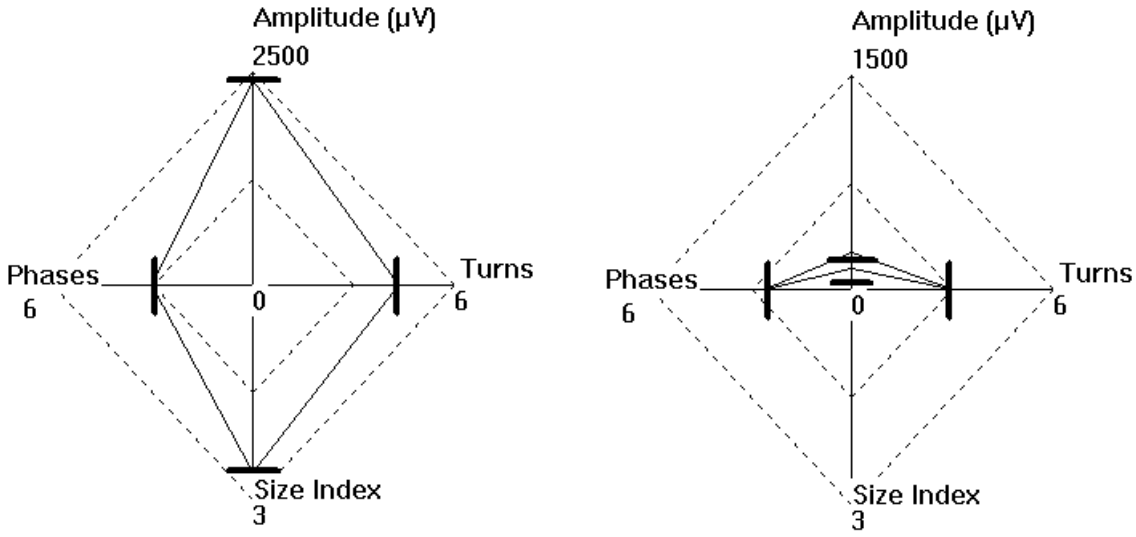
Figure 4-14 The 5-dimensional polar plot for a neuropathic subject

Cross Graph

A reduced polar graph to be called a cross graph was experimentally developed in an attempt to create more unique patterns to distinguish neuropathies and myopathies. The cross graph has only four axes and the three major categories of condition each create a type of diamond shape upon the graph. The normal diamond is approximately a square, turned 45 degrees to rest on one corner. It was considered that by making some axes represent the inverse of some properties one could make a graph that narrows in one direction for myopathies and in the other direction for neuropathies. The concern there would be that the physician could be confused by an inverse function, especially if it regards an already unfamiliar characteristic. For this reason, inverse axes were not used in any displays.

The present DQEMG cross graph design presents peak to peak voltage (amplitude) in the upright vertical direction, phases and turns to the left and right, and size index in the downward pointing vertical direction. A Neuropathic case will present larger in the vertical direction and show a stronger effect in turns than in phases (Pfiefer and Kunz (1992), as cited by Zalewska and Hausmanowa-Petrusewicz (2000)). Neuropathic data will create a large diamond shape on the

cross. Putting the phases and turns in opposite directions on the same axes is expected to make comparisons between them easy. A Myopathic case will be smaller in the vertical direction than the horizontal. It may not be distinguished from a normal case by either turns or phases, but it will appear to be a narrow horizontal shape that will be easily recognizable in some cases; the size index mean might be below zero, which would collapse the bottom of the diamond and create a triangle.



a. b.

Figure 4-15 Cross graph displays for a) neuropathic and b) myopathic cases.

4.6.2.4. An Integrated Graphical Display

For the purpose of evaluating the practicality of these graphical display options, they were combined into a prototype graphical display in DQEMG, shown in Figure 4-17, on the following page. This display had five histograms (amplitude, duration, firing rate, size index and thickness (AAR)), one scatter plot of amplitude and AAR, a 5-dimensional polar star and a 4-dimensional cross plot, as well as two additional statistics that were not calculated for the Muscle Study Results screen. One goal was to get feedback on the new statistics, % polyphasic MUAPs and FR/MU. The percent of motor units in the study that were polyphasic, or had more than 4 phases, was reported in yellow text near the polar stars. The FR/MU was also calculated and displayed at the bottom of the graphics screen, just below the firing rate histogram.

A button for navigating to the graphics display was added to the bottom toolbar in the Results screen. The final results button bar used in the Expert testing is shown below (Figure 4-16).

Figure 4-16 Bottom Results toolbar with Graphs button

The graphics display was designed as a subscreen of Results, so the user could return from it to Results using a **Close** button, consistent with other subscreens.

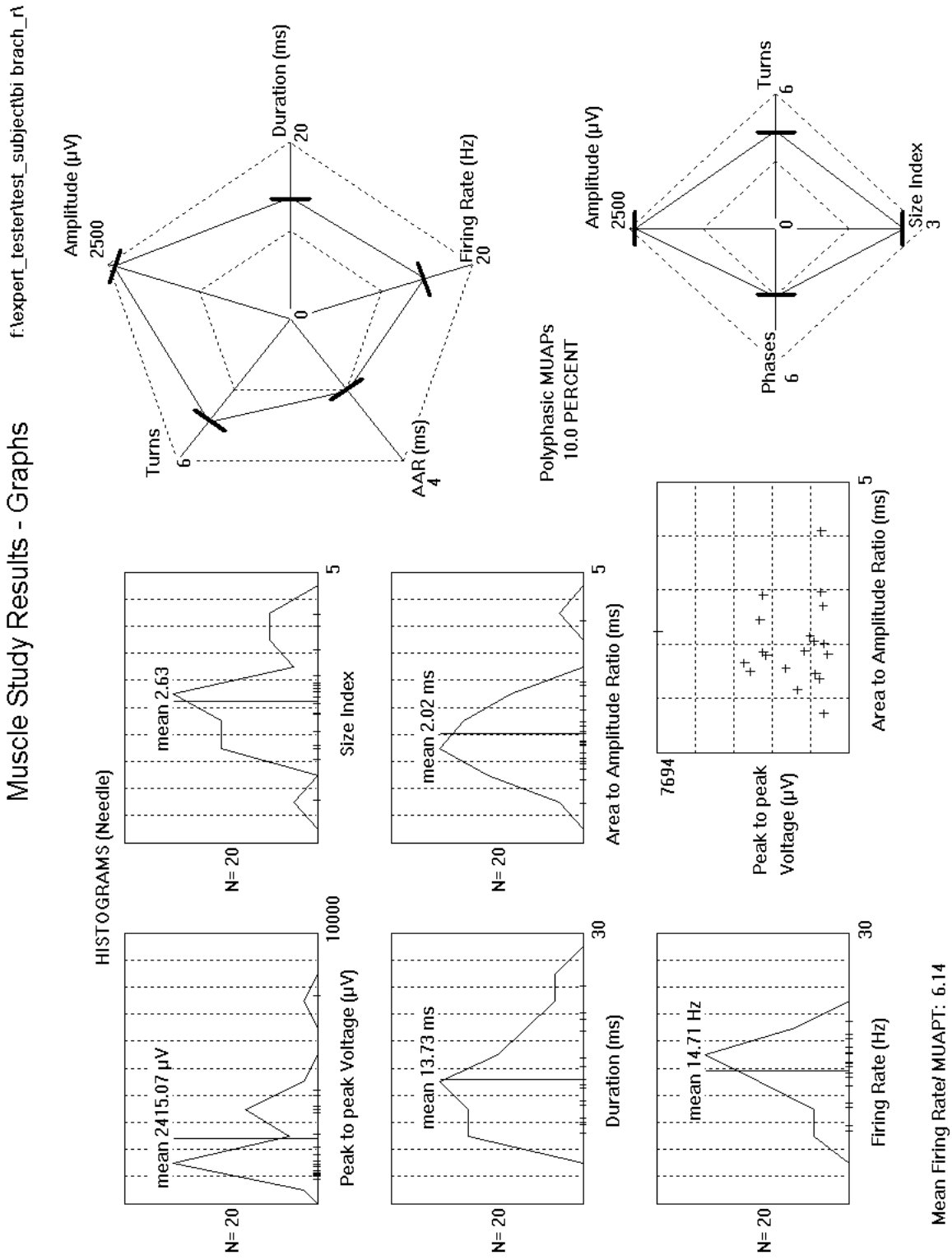


Figure 4-17 Prototype graphical display (archived neuropathic data used in Expert testing).

Chapter 5 Expert User Testing Methods

5.1. Objectives

This research explored the user experience using the decomposition-based Quantitative Electromyography (DQEMG) application. The DQEMG application gives the user quantitative information about EMG data they have collected from a patient. The purpose of the research was to determine what information from the application is most useful to the user in aiding the diagnosis of neuromuscular disease (or health). In addition, observations about workflow in the application and errors (both navigational and functional) made by users will provide the basis for improving the usability of the DQEMG interface.

The DQEMG application was developed for use in clinical situations, by physicians. Since their time is both short and costly, the chances that DQEMG will actually be adopted for use will be higher the more quickly and efficiently they can use the application and the more helpfully presented the information is. The first hypothesis was that there are certain areas where the application is awkward or time-consuming to use, which would be identified in this study and could then be improved appropriately. These were expected to correlate with the issues identified by the initial interface analysis. The second hypothesis was that the current design may not do a good job of isolating and presenting the information that would best improve the ease and effectiveness of diagnosis by using DQEMG.

A redesigned graphical information display was included in the presentation to the experts tested, and comments were solicited as to its usefulness. This is a type of expert user assessment that is often used on its own. See the *Design Suggestions* portion of Chapter 4 for discussion of the design of the prototype display components used in this graphical display.

A concurrent verbal protocol was chosen because it can capture information about how a user navigates through an interface (Ericsson and Simon, 1984). The participants spoke aloud their thoughts and actions while doing tasks on the system and their words were recorded. A video recording of the computer screen was also made to capture the user's actions. The user was prompted to keep talking (to "think aloud") if they fell silent. This is a somewhat intrusive protocol, but Ericsson, Simon and others have demonstrated that the use of a concurrent verbal protocol does not have a significant impact on task performance if it is done carefully. While users tend to verbalize mainly what they are doing and not why, this protocol was anticipated to capture indications of how the user was feeling, whether or not they were lost or frustrated, and also statements about what information they were using from the interface that was causing their

navigational decisions or errors. A verbal protocol is known to do a good job of capturing the information in short-term memory that the user is attending to (Ericsson and Simon, 1984).

Because the landmark marker editing task in the application had been identified as a problem spot both through casual user comments and through the HCI interface analysis, a couple of additional performance measures were collected regarding this portion of the user testing. The time it took the user to complete editing markers in 20 MUs was measured and recorded, and a NASA TLX assessment was chosen to collect information from the user regarding the cognitive and emotional load and reaction they experienced during that task (Hart and Staveland, 1988). While a qualitative assessment of the TLX results may point to particular information to support the results of this study, a quantitative statistical analysis would be unreasonable considering the sample size. However, gathering the data in this study will permit a future quantitative comparison between the user experience on the current interface and the user experience on a resulting redesign, which would hope to have improved that experience. The assumption behind this reasoning is that the Biological Signal Detection and Analysis Lab will hold additional user testing at some future time. Improving the landmark marker editing task is expected to improve the speed with which a user can use the application, and at the same time reduce user frustration.

It is hoped that this study will result in a more effective and usable interface for the DQEMG application, raising the chances that it can be successfully introduced into the regular process of EMG diagnosis. It is believed that the addition of Quantitative EMG into the normal operations of EMG physicians will improve the efficiency and effectiveness of EMG-based diagnosis, treatment and management of neuromuscular diseases. This would provide a clear benefit to society. Results from this study are also anticipated to benefit the scientific community working on Quantitative EMG.

5.2. Purpose and Design

The purpose of the project was to observe how trained EMG professionals may use this software, and which parts of the software either impede or support the ultimate goal of the program, which is to aid and improve the clinical diagnosis of neuromuscular disease. The basic method then was to place each subject in a simulated situation much like their professional one.

In order to familiarize the participants with the muscle study data acquisition and analysis procedure using the DQEMG application, the participants were first asked to complete an MVC protocol and collect data from one contraction from a volunteer “patient” (a normal person in fine health). Using the practice muscle study from the healthy “patient”, the participant was coached to interact with the DQEMG interface in such a way as to learn the basics of how it was organized

and what they would need to do in order to assess a muscle study. Participants were encouraged to ask questions during this phase of the experiment. One of the six participants was unable to collect a contraction due to a persistent computer error. That participant was coached using an archived muscle study from a previous study subject.

After the practice study, the participants were asked to approach an archived muscle study given a scenario of a patient who reported a problem with their biceps brachii, the large muscle in the upper arm that causes the elbow to bend. Specifically, they were told the patient had been clinically diagnosed with a right C6 radiculopathy and complained of biceps weakness (the complete script for this is in Appendix A). C6 is the spinal nerve group that innervates the biceps and other muscles in the arms. A radiculopathy is a type of nerve disease, or neuropathy. The participants were asked to process and examine data that had been analyzed by our system to check for involvement of the biceps brachii, with a goal in mind of characterizing the results of the EMG study. During this procedure a verbal protocol was recorded from the subject, and a video recording was taken of their actions on the computer screen. Before and after this procedure they were given some explanation about the verbal protocol and the procedure. They also filled out questionnaires regarding their experience with computers, and their thoughts relating to the DQEMG application based on any previous experience they had with it. We also administered a NASA Task Loading Index (TLX) immediately following the markers editing task, something we expected them to have problems with. This gave us an index with which to evaluate future improvements to the system.

The markers editing task involves editing the landmarks of the MUAP templates. This task was also timed by the researcher, and notes were taken throughout the test and during later analysis of the video regarding observable errors that each user made. Through the use of a standardized data set for the landmark editing, the participants were presented with editing that ought to have been done a certain way, so not only could the usability of the interface be assessed, but also whether it lead the user to correct or necessary actions or perhaps impeded those actions from happening. Unfortunately regular system crashes impeded our ability to save the specific changes the users made to the landmark positions in enough cases that this particular direction of analysis was abandoned.

5.3. Participants

The Expert study involved 6 professionals trained and experienced in electrodiagnosis, recruited from the population of trained EMG technicians and doctors within 300 miles of London, Ontario.

Participants were recruited through existing contacts in London and mainly through verbal or email dialogue with our affiliated physicians. Participants' ages ranged from 30 to 51. The mean age was 38. All the participants were male. This gender imbalance was due to the available pool of these professionals in this region. All but one of the participants had a medical degree (MD); that participant had a BA. Of the other participants, two had Bachelor's degrees, one had a Ph.D., and all five had additional specialization training; three in physical medicine and rehabilitation (with one of those also having specialization training in EMG), one in Neurophysiology, and the other in Neurology.

Two of the participants had previous experience using DEMG. Three of the participants reported they use a personal computer about 5-9 hours per week, which is considered minimal. One participant used a computer 10-19 hours weekly, while two reportedly spent 20-39 hours on a computer each week.

Participants were not compensated in any way for their participation. All subjects gave informed consent to participate in the study. This study took place off-campus at the Saint Joseph's Hospital, part of the London Health Sciences Center.

5.4. Apparatus and Materials

Neurosoft's Comperio system was used to collect a normal EMG signal and test the data interface for the procedure of importing data into the DQEMG application.[2] The Comperio is an electromyograph, which displays the EMG signal on a computer screen, plays it as a sound, and permits the physician to alter the volume of the speaker and the scale on the screen. There is a function key in the Comperio interface that leads to the DQEMG application. The user can either click on a button on the screen or use a button on the keyboard to invoke the DQEMG "analyze" function.

A standardized set of data was used in the analysis task. This was stored on the computer and opened in the DQEMG application at the appropriate time. The data was real data from a neuropathic patient that was deemed appropriately representative for the task scenario. The archived study was edited to have seven contractions in it, and it was considered to clearly indicate a chronic neuropathy.

[2] Neurosoft /Neuroscan is located in El Paso, Texas.

Figure 5-1 Equipment for Expert testing: the Comperio system and video camera.

As Figure5-1 shows, the Comperio system is conveniently mounted on a multilevel cart. The monitor and a regular keyboard are on the top of the cart, leaving room for the physician to prepare for the EMG protocol by keeping gauze, medical tape, and needle electrodes to the side if necessary. The second level is a shelf that can slide out to give the user access to the Comperio controller, which includes a specialized mouse ball and other controllers and knobs. On the bottom level is the CPU. The CPU pictured above was stolen the night before the 5th participant's session, so the last two participants ran the protocol on the system as shown but with a laptop on a chair to the right of the Comperio cart. The laptop display was output to the large monitor so the system worked more or less as normal though I/O limitations on the laptop meant the regular sized keyboard could not be used (the specialized Comperio controller was connected to the laptop's keyboard input connection on the laptop, so the laptop keyboard had to be used).

A Sony ICD-BP100 digital voice recorder was placed to the right of the Computer monitor to capture the verbal protocol. A digital video camera on a tripod recorded activities on the computer screen during the procedure and provided a backup audio recording. The researcher took notes throughout the testing and used a stopwatch to time the marker editing and study characterization processes. Various questionnaires were also applied. See Appendix A for written materials used to collect data for this procedure.

5.5. Procedure

The protocol for the design evaluation called for standardized questionnaires distributed in person, and computer-administered tasks, some of which were standardized. Audio recordings of the participants' verbal protocol and video recordings of actions on the computer screen were taken, as well as unobtrusive observations during the testing. A script was used in order to standardize instructions given to the participants as much as possible. The complete script is attached as Appendix A-1.

An information letter and a consent form were presented to the test participant. Participants were informed that no risks are anticipated from participating in this research. They were presented with the scenario of the task and were not deceived into thinking the task itself had any impact on anyone. They were informed that they were free to withdraw at any time. Explaining the basis of the study, and informing the participant of the nature of it, was anticipated to protect against possible stresses to the participant, and designed to support an ethical research procedure.

An Entrance questionnaire was administered to gather demographic information and allow us to check for confounding variables. Level of regular computer use, training and education, age, and previous experience with the DQEMG application were identified as the most probable confounding variables, so information was collected about them. The average number of hours spent on a personal computer each week is assumed to indicate level of familiarity and skill with computer interfaces. In general, if a highly trained, very computer-active person was to get confused by our interface that might be a sign something is really wrong. It may be that our interface has failed to follow convention or that a needed action is hidden in the interface. If a person with less developed comfort with computers had trouble, this would tell us something else. Perhaps the visual cues for interaction are not clear enough or there is not enough direction to the workflow. Since our subjects were physicians, we expected a significant amount of their regular daytime tasks would *not* be done in front of computers. Part of the reason this question was included was to see if that assumption was correct.

It was also considered important to record age so that other usability tests done in this area can be compared to this evaluation, and to demonstrate that the sample group was representative of the target population. Additional questions asked the participants to describe their clinical goal in the given scenario and to indicate what tasks they would be expecting to do with DQEMG, what electrophysiologic evidence they would be looking for. This would identify participant's biases and expectations regarding what useful information or data they might seek or find in a needle EMG study. Finally, those participants who had used the DQEMG application before were

asked to comment on their previous experience with it and list things they particularly liked or disliked about it. This was a memory-retrieval prompt. Anything solicited by these questions was expected to be very memorable.

The DQEMG application and how to use it was introduced with a two-page written overview (see Appendix A-4). Verbal Protocol was explained and the subject was given an opportunity to practice it while doing a task on the Comperio System (entering patient and muscle information for the study they were about to do). The participant was given feedback on their verbal protocol practice. The audio- and video-taping began at this point. The computer screen was video taped, but the participant was not. To facilitate the whole testing process and give an overview to the user, a Procedure Card was taped to the left side of the computer monitor (see Appendix A-5).

Two other questionnaires were administered. A NASA TLX (Appendix A-6) was filled out right after the landmark editing task, which was a problematic task of particular interest. The TLX was developed as a way to consistently measure different aspects of effort expended on a particular task. Participants filling out the NASA TLX indicate their subjective sense of mental demand, physical demand and temporal demand (how hurried and rushed the pace of the task) along a 20-division (unnumbered) scale from Very Low to Very High. They also report their level of success or task performance, from Perfect to Failure, along with how hard they had to work to achieve that level of performance (effort) and how frustrated, insecure or annoyed they were (frustration), on a scale from Very Low to Very High (Hart and Staveland, 1988).

An exit questionnaire was filled out at the end of the session. This questionnaire was designed to record the participants' characterizations of the EMG pattern in the standardized data, as well as comments on which parts of the interface helped or hindered the participant in coming to that conclusion.

After the first user test it was determined to be necessary to review all handwritten documents to check for legibility. This was done at the end of each user session after that; clarifications from the participant were recorded by audio recording as well as by hand.

Observations were taken as to errors in the task executed by the participants. Examination of the videotapes produced identification of more error points and helped give context to others. Inappropriate action given the participant's goal, omission of important steps, repeated or reversed navigation, and errors that the participant commented on were all considered errors in the course of this study. A list of terms and phrases like "oops" and "I didn't want/mean to do that/go there" was compiled beforehand to improve the consistency of verbal protocol analysis.

Forms and recordings were identified by a participant number rather than a name. Data saved electronically was filed according to these numbers so as to protect the anonymity of the human subjects. Audio and video tapes were also identified by this number system.

5.6. Measures and Variables

With such a small sample size, we did not vary any independent variables. All the users were in one group, presented with approximately the same scenario, equipment, and data. The data in the first real-time data collection portion of the exercise did vary but as it was all taken from normal subjects it was not expected to vary to any great degree, and since they were mainly analyzing the standardized archive data, variation in the collected data was not expected to have an impact on how well we can generalize our observations. The equipment varied more than had been planned due to the system CPU's having been stolen mid-process.

Dependent variables were: the time taken to do the markers editing task, the total time spent in the data analysis, the number of errors in navigation, the number of oversights or erroneous corrections in the editing task, the subjective reporting on landmark editing given in the NASA TLX profiles, and counts of different types of navigational, observational and emotional comments made in the verbal protocols. Exit questionnaire comments were also categorized and summated, though again the sample size is too small to describe this as a statistical analysis.

Chapter 6 Expert User Testing Results

6.1. Entrance Comments

Entrance Questionnaire responses, NASA TLX responses, and Exit Questionnaire responses are collected in Appendix B. The two participants who had previously used DQEMG reported liking how the program appeared to collect and analyze data quickly, the extent of MU identification and decomposition accuracy, and the thoroughness with which waveform templates could be edited. Both reported disliking the amount of time required for editing.

In describing their clinical goal with the given scenario, two participants described seeking to assess the presence or severity of axonal damage or loss and another mentioned relating the physical exam to the clinical EMG. Three participants referred to ruling out other possibilities in addition to considering the proposed diagnosis [C6 neuropathy]. Regarding what evidence they would look for to determine the presence of a C6 neuropathy, all six participants said they would look for reduced recruitment. Four participants mentioned they would look for spontaneous activity; fibrillation potentials and positive sharp waves. Three said they would look for polyphasicity and three mentioned MUAP size. One mentioned increased firing rates and another decreased firing rates. The latter participant was in error in that firing rates are increased in neuropathy, but it's not clear whether the participant had a mistaken understanding or simply slipped when writing down his thoughts. Two participants hoped to do motor unit number estimation. As expected, none of the participants mentioned editing MUAPs or assessing the decomposition in the list of tasks they expected to do with DQEMG, even thought two participants were familiar with the application and another had seen it used.

6.2. Correlations

Though there was a small sample population for this study, some of the numerical results had sufficiently large correlation coefficients for their correlation to be considered statistically significant. (However, all discussion of numerical patterns in NASA TLX results should be viewed lightly, as discussed below with regard to relative numbers that are reported on a subjective scale. The patterns are interesting, but the NASA TLX is mainly for comparing results within the same participant on multiple tasks (Hart and Staveland, 1988).) The threshold for N=5 with N-2 degrees of freedom for significance at the 95th percentile is $t = 2.353$, or $\rho=0.81$

(Mendenhall and Reinmuth, 1978). For N=6 the ρ threshold was only 0.73 and there were six participants, however participant 1 was short on time and reviewed only some of the templates, editing none in the 2 min 33 seconds he gave to the task. It seemed he had essentially skipped the task, so his NASA TLX responses were not considered to have the same context as those of the other participants, and therefore participant 1 responses were omitted from the statistical analysis of the numerical responses and N was reduced to 5 for statistical analysis having to do with that task.

The correlation coefficients for all the numerical responses are in Appendix B-2. There was a positive correlation between regular computer hours and reported sense of mental demand (ρ=0.84) involved in the markers editing task. Stronger correlations were found between regular computer hours and reported effort (ρ=0.9) and frustration (ρ=0.94) on the same task. It was considered possible that the more experienced computer users spent more time editing the markers, leading to more frustration, but in fact, there was no significant correlation between the amount of time taken and reported frustration. The only reported characteristic that had a significant correlation with time was the age of the participant. There was also a positive correlation between the amount of training the participant had and the level of temporal demand they reported, and there were correlations between reported mental demand and effort and between frustration and physical demand. It's possible that some people just tended to report things higher on the task index (TLX) scales than other people, which would lead to the observed positive correlations between TLX factors.

One hypothesis of this research was that if a participant with a lot of regular computer experience had trouble with the interface, it might not be following established interface design standards well enough to fit the participants' mental models of how applications work. The reported levels of effort and frustration on the part of our experienced computer users (participants 3 and 4) therefore raise concern in this area. The next step will be to see if they showed signs of confusion or experienced error during their interaction with the application.

6.3. Errors and Confusion, Comments, Complements and Suggestions

The verbal protocol was transcribed and the video recordings were analyzed to identify five types of event in this user testing: Errors, Confusion, Comments, Complements and Suggestions. A complete list of those events, organized by topic, is included in Appendix B. Categories of particular interest due to their implications for the DQEMG interface will be discussed here.

6.3.1. MVC Feedback Needed

Three participants (1, 3, and 6) expressed confusion when faced with the blank DQEMG screen following the MVC protocol. Some of them made multiple comments regarding not knowing what they should do next. The procedure card says "Click **Acquire** to return to the EMG screen and select channel 1" under the next step, but the participants were not sure the MVC step had been completed or that the MVC had been collected successfully. Two participants (1, 4, and 6) made comments indicating that they'd like feedback that lets them assess the quality of the MVC collected. One participant (3) concluded he should hit the "open" button next to open "it" (perhaps he thought he could open the MVC? This was not clear).

6.3.2. Modes – Edit, Markers, Details and Raster

Four participants expressed a desire to edit markers from the Decomp Screen combined with confusion about how to do so. One participant wasn't sure how to do it even after he had been instructed through it, had done marker editing for a while, and returned to the Decomp screen again. Another participant (3) asked how to edit the markers from the Needle screen. Three participants tried to edit a Micro template just by clicking on it. One asked if it required a double-click. One participant repeatedly left the Decomp screen to edit markers in the following fashion: first he would hit the **Markers** button in the Decomp screen. Then he hit the **Needle** button on the top toolbar and switched to the Needle screen, then he finally opened a template by hitting the **Markers** button in the Needle screen and selecting a template graph. He never selected a template graph directly from the Decomp screen.

One participant (3) saw the **Edit** button while trying to edit the markers, selected it, and then selected a template, clicking on it a second time when it didn't open. Three participants attempted to select a template that was gray when they wanted to edit the markers on it, not understanding that the color gray indicated it had been edited out. A fourth participant commented that it would be nice if there was an explanation on the screens for what the colors yellow and gray meant. Two participants who were marker editing by selecting templates from the Decomp screen lost track of where in the list of MUAPTs they were and weren't sure which template to open next. One participant suggested there be a checkbox or something to let you know what you've already looked at.

Participants who tried looking at the shimmer plots had similar difficulties. Two participants had difficulty just figuring out how to open the shimmer plots. One started out

thinking the firing rate information was what was meant by "shimmer plot" (he apparently missed seeing the column headers) and another never did open the shimmer plots, having failed to figure out how. Participants who did use both the Details and Raster functions got confused between the two of them, and could not remember which one to select if they wanted to be able to edit the shimmer plot.

A few errors stemmed from the fact that a mode continued to be active and participants would lose track of what mode they were in. One participant tried to click on a shimmer plot while in markers mode. The mouse click selected the macro template to the right of the shimmer plot, which was the most recent template the mouse had passed over. The participant recognized the error and backtracked to look at the shimmer plot. A couple of participants clicked on templates to change the markers while in edit mode. One of those participants had switched to a new contraction and he commented that the program should not remain in edit mode after that kind of navigation. One participant tried to turn the edit mode off by clicking on the **Edit** button while the edit mode was on, treating it like an on-off switch.

6.3.3. Shimmer Plot

Participants were uncertain how to evaluate the shimmer plot both on the Decomp screen and on the Details and Raster screens. On the subscreens in particular they expressed confusion about what makes a good shimmer plot, and asked for the numbers on the screen to be labeled more clearly. One participant thought it might be easier to see the waveform of interest without all the noise of the signal to either side of it. A couple of participants complemented the information displayed in the shimmer plot. One of them felt it would be most useful to look at the raster plot while deciding where to place the onset and offset landmarks on the MUAP template (especially macro MUAPs), and suggested it should be possible to access the shimmer graph directly from the marker editing screen. One participant commented while looking at shimmer plots that it was a slow process, and took patience. The couple of times that participants began to make edits to the waveforms in the raster plot (i.e. editing specific waveforms out of a specific MUAPT), Dr. Stashuk explained why they ought not do that. These repeated explanations lead our design team to realize that while the editing function was perhaps useful for research purposes it should be removed from the clinical release of the application.

6.3.4. Navigating through Contractions and MUAPs

The participants all had or expressed some kind of difficulty navigating the collection of contractions. The first thing a couple of participants commented on was that they didn't know how many contractions were in the study. When they realized there was no *next* contraction, and they had to either go backwards or go all the way back to the beginning, most went to the beginning and then proceeded forward through the study. A couple edited the MUAPTs as they went backward and then edited the markers in a forward direction. One of those participants got confused going through MUAP templates and could not remember if he had just done *previous* or *next* and had to check the previous MUAP to confirm he had already edited it. Another participant made an error going backward and hit **Next** instead of **Prev**. Two participants commented on the way the **Next** button being gray means they're at the end of the contractions. One of them clicked on the grayed-out **Next** button first before realizing it wouldn't work. Finally, one participant never did realize the marker editing screen had **Next** and **Prev** buttons and closed the editing screen between every MUAP. He commented that it would be nice to be able to edit them in a series without having to do that. Similarly, looking at the Details (shimmer plot) screen where the **Next** button takes the user to the next page of waveforms from a single MUAPT, a participant commented it would be nice to be able to switch from MU to MU without having to close the screen and go back to Decomp. Participants also wanted a way to move to other contractions while in the marker editing screen.

6.3.5. Marker Editing

Although there was sometimes initial confusion about how to move the markers on the marker editing screen, most participants mastered it quickly and three participants commented that it was easy or not hard. One participant said that the numbers were confusing and two participants tried to click on the text labels in the key in order to select the markers before finding out they could click on them directly. Two participants also tried to select the red marker that indicated a macro landmark. One participant complained that it was a little tricky selecting the markers with the mouse, that you had to be right over them. Another participant complained that when two were close together it was hard to pick them up individually. When the bottom toolbar buttons were explained to him, he complemented that idea, but did not use them.

Since the archived study was from a neuropathic subject, all of the MUAPs had peaks that went off the editing area at the default scale. One participant commented that it would be nice if the whole thing appeared within the box so you wouldn't have to change the scale all the time.

Three other participants complained about having to change the scales repeatedly, and one suggested that it should be possible to change the scale for the whole set of MUAPs with one action. Two participants did note that the peak landmarks were usually fine, and it was mainly the offset marker and sometimes the onset that had to be changed, but most participants wanted to look at the peaks, even if they didn't expect to have to change them. One participant complemented how a landmark button's being grayed out gave an additional clue that the landmark was off the screen.

All of the participants expressed uncertainty at some point about where to place a marker. One complained that there was too much guesswork on onset and offset placement, and also complained that the gray color used for the macro signal on the Micro screen was too hard to see. Two participants were observed picking up a marker and putting it back down in the exact same place. There were also a couple of errors observed, where the location a marker was placed at was clearly wrong. One participant commented while filling out the NASA TLX, "I don't know, how successful was I? I have no idea." Uncertainty and slow decision-making combined with this type of micro-editing – moving landmarks a very small distance or even moving a marker and then moving it back to where its original position – could be part of the reason marker editing takes a lot of time. One participant complained repeatedly during the marker editing about how long it was taking.

6.3.6. Scale and Sweep Buttons

There were a number of errors changing the sweep and the scale in various screens. Four of the errors involved hitting the wrong sweep or scale button and having to reverse the effect. One of the participants who made an error like that was comparatively highly experienced both with computers in general and the DQEMG program in particular. In another case, a participant hit the sweep button when he meant to hit the scale button. One participant suggested that for changing the scale so often it would be helpful to have keys to do that at his fingertips –"up-down, up-down, right there."

A couple of participants had difficulty locating the scale buttons in the first place and had to ask how to change the displayed amplitude. One participant commented on the Decomp screen that the buttons could be labeled better. He suggested using up and down arrows instead of the current icons.

6.3.7. Closing Subscreens and Switching Screens

Five out of six participants had problems getting out of the subscreens and back to where they had been previously. Some participants tried using the top toolbar when they failed to find a "Back" button and were frustrated that the button for the parent screen was unavailable. One participant complained that "Close" puts him in mind of closing a document, not a screen, and one participant hovered the mouse around the **Close** button in the top toolbar (*Close study*) long enough to read the mouseover and realize it wasn't what he wanted. One participant even verbalized that he wanted to "close this" but failed to locate the **Close** button and hit **Summary** instead when trying to return to the Results screen from Graphics. Another participant who was on the marker editing screen hit the **Needle** button and then asked how to get back to Decomp.

Some confusion probably stemmed from the fact that these were mostly novice users who hadn't learned the names on the screens and the buttons that would lead to them. One participant hit the **Ensemble** button when trying to get to "the main screen" which turned out to be the Decomp screen. Four participants expressed confusion about what "results" and "summary" meant and two participants lost track of which of those two screens they had been on. One said "Let's go back to Results" when he hadn't been on results yet, he had been on Summary.

The participant with the most experience with the DQEMG application suggested there should always be a big button that does the next action the user is most likely to take, e.g. "Return to Decomp" when at the last Micro template in the contraction. He also put a comment on the Exit questionnaire about getting lost between screens and said that it wasn't always clear how to move backward or forward.

6.4. Useful Information

The expert testing participants commonly identified the micro MUAP statistics as useful in both their comments and their exit questionnaire responses. Amplitude, duration, phases and turns were all singled out, especially amplitude. Size index, % polyphasic and %MVC were also complemented. There were a couple of requests that ranges of values be listed in addition to (or in place of) standard deviations. One participant included area-to-amplitude ratio in a list of MUAP statistics that were good, but the rest of the participants said they didn't know the numbers for AAR because they don't use it, and one participant demonstrated confusion about the definition of thickness and didn't know what AAR was at all.

One participant was familiar with macro EMG amplitudes and two indicated that macro EMG was useful for MU number estimation (MUNE), but most participants were not familiar with how to interpret macro information and could make no use of it.

Firing rate was identified as useful, but two participants were unsure how to interpret it and one in particular complained multiple times that it didn't make sense to average firing rates together. He did complement the way firing rate information was listed for each individual MUAPT on the Decomp screen, but didn't know how to interpret the Results display. He was looking for a more definite way to know about recruitment for the level of contraction. Another participant said firing rates needed to be interpreted in light of the force level and wasn't sure he could get that from %MVC. Participants generally weren't sure what IDI was and one participant suggested leaving IDI off the results list, complaining that the screen was already too visually busy. The same person suggested that identification rate probably doesn't mean anything to an electromyographer and shouldn't be reported in study results either.

Size index and FR/MU were unfamiliar to most of the participants, a situation which was expected; only one participant knew that a normal FR/MU value was supposed to be around 5. The polar star graphs also had to be explained. At first glance, one participant expected something on the graph to indicate one or two standard deviations of normal (a suggestion executed in the design for the Nonexpert testing). Four participants listed graphs and figures as useful information in the exit questionnaire, and comments on seeing and using them were generally very positive (e.g. "This is cool.").

Most of the participants were not familiar with quantitative norms and 5 out of 6 of them commented about needing or wanting them in order to interpret the data. Participant 2 suggested that the best way for the application to be structured would be to provide a way for each lab to collect its own normative data in a database. He commented that equipment and procedures vary from lab to lab, so normative values might vary in the same way. It is also apparent that physicians would probably have more confidence in norms they had established themselves.

Three of the participants complemented the data presented in the Results screen.

6.5. Time Taken for Marker Editing and Decision-Making

Participants were timed from when they stated they were beginning the marker landmark editing to when they said they were done with it. They were watched carefully and in some cases they were asked if they were finished if it seemed that they were. The marker editing times varied from 2 min, 33 s to 26 min with an average of 14 min, 30 s. As was mentioned above, the participant

who only took two and a half minutes seemed to essentially skip the task; when he is removed from the group the average time rises to l6 min, 53 s. It may be more meaningful to examine groups of times than the mean, however. Three participants took between 8 and 14 minutes to complete the marker editing, while two participants took nearly twice that – 23 ½ to 26 minutes. As was mentioned above, there was a correlation between age and marker editing time. The longest time was taken by the eldest participant, who was 51. All the participants in their 30s were in the faster group. The group that took longer did more micro-editing. Another difference between groups was the amount of interest they took in the macro markers.

It was difficult to record a consistently defined time for decision-making, but basically the time recorded was how long the participant took between when marker editing was done and when the participant was willing to characterize the study. Participant 2 has been omitted, since he spent a lot of extra time aimlessly wandering through the program micro editing and commenting on things, and it did not seem his decision-making objective and procedure was comparable to any of the other participants' procedures. Decision-making times ranged from 11 min to 28 min, 51 s with an average of 18 min, 54 s. This was not including some extra time two participants spent reviewing graphical results after they were prompted to do so. Including that time makes the high end of the range 34 min, 9 s and brings the average up to 20 min, 39 s. The timer was halted during system crashes or technical problems; they did not affect results.

6.6. Characterizations

Five out of six participants reported a characterization of the EMG study during the exit questionnaire. The other participant characterized it as “Detailed” and may have misinterpreted the question, but made no comments indicating he thought it was anything but normal. Two of the participants characterized the data in the study as representing a chronic axonal process, based on the large MUAPs (increased amplitude with long duration), and lack of significant polyphasic activity. One participant characterized it as a Neurogenic condition, based on the C6 presentation [scenario]. That participant also mentioned larger motor units and high firing rates. Two of the participants classified the archived study as normal. One of those participants commented that he had seen a few very large MUAPs in the first contraction but the rest seemed normal and perhaps there was some sampling error? In fact, the smallest peak-to-peak voltage in the archived study is 896.6 μV, with the next smallest above 1000 μV. That’s well over two standard deviations above the normal mean, so none of the MUAPs the participants saw would typically be described as normal. One might expect to see one or two MUAPs that size in a normal study, in a normal case

they would not represent the low end of MUAP size. The other participant who characterized the study as normal noted that the size index values indicated otherwise, but was not confident of how to interpret them.

Five of the participants commented that they needed or wanted normative data as a context in order to be able to properly interpret the information. One participant reported that he usually used monopolar needles, so concentric amplitudes and such were unfamiliar to him. One participant who characterized the archived study as normal commented both during the session and on the exit questionnaire that the physical exam was the most important contributor to diagnosis.

6.7. Confidence and Significance

The exit questionnaire also collected a report on the confidence each participant had in the information from the DQEMG application and what significance they attributed to the contribution of the quantitative information to their conclusions about the study. The reported confidence in the DQEMG information was moderate for all participants, without much variation (the mean was 3.75 +/- 0.4 out of a possible 5 points on a scale from not very confident to very confident). All but one participant marked a lower position on the scale for the next question, "Do you feel the quantified statistical information provided by the DQEMG application contributed to your conclusion?" There was a significant correlation between reported confidence and the level of effort the same participant had reported on the NASA TLX, and also a correlation between how important they rated the contribution of statistical information to their decision and how high they reported the level of mental effort for the marker editing task. So it seems that if they tried harder and believed they thought hard about the marker editing task, the participants were more likely to have confidence in the DQEMG evidence and give weight to it in decision-making.

6.8. Technical Problems

In addition to the user activities and reactions that were noted during the testing, we also tracked how many times the EMG EP and DQEMG applications failed to complete an operation, gave an error message, or crashed. This happened more than expected, and was of understandable concern to the participants, so it seems appropriate to report it here in order to give context to their comments.

There were at least five program crashes during the testing, mostly during signal acquisition, with two crashes occurring later in the run of the program. This happened once to participant 2, twice to participants 4 and 5. Three times the crash involved the error, "This program has now performed an illegal operation and will be shut down." Another time (with participant 2) the program had been running a long time and when the user returned to the Graphics screen it was blank below the title and the system would no longer respond. One crash involved bizarre "ghosting" behavior on screen where various windows opened and functions were executed, seemingly by a ghost, and none of the user's actions had any effect; to restart after this crash the participant had to reboot the machine. Participant 6 experienced two "Get data failed: Buffer overflowed" error messages, and one oddly redrawn screen where the window was cut off on the right side, but nothing that made the program crash. The only error participant 1 experienced was in trying to use the Compare to Database function from the Results screen, which did not work. Participant 3 is the only participant who completed the session without application error.

Obviously system crashes are not a designed *feature* of the program; these difficulties are a symptom of stage of development that DQEMG was in during testing, and also some ongoing problems with the EMG/EP application from Neurosoft. Most of the technical problems described here have been resolved since the testing.

6.9. Willingness to use DQEMG in Clinical Practice

At the end of the Expert testing sessions three out of six participants said yes, they would use DQEMG in clinical practice. One said he would only use it for MUNE and might use it for regular EMG but it would depend on time vs. improved diagnostic yield. One participant answered “Not sure”. He and one of the participants who had said “yes” qualified their answers with concerns about the reliability and validity of the decomposition and wanted evidence it would change the diagnosis and treatment. The participant with the most DQEMG experience said yes and no –yes, because it provides useful quantitative data, no because it crashed three times during his session. Another participant concurred that a physician does not have time to repeat procedures if an application crashes. With the application now more stable, we can presume the participants would use it clinically, given a demonstration of such. The results of rigorous functionality testing according to a set testing protocol in a clinical setting should be sufficient to demonstrate that stability.

Chapter 7 Nonexpert Testing Methods

7.1. Objectives

The purpose of this research was to determine what quantitative EMG information is most useful to the user in aiding the diagnosis of neuromuscular disease (or health). The experiment used a non-computer task that examined perception and decision-making based on information gathered in a single display printed on paper.

The rational of this project was to comparison test three modes of displaying the same information to a person who is trying to discern the state of a patient based on several reported pieces of data. This is a within-subject repeated measures experiment. Specific measures of efficiency and effectiveness were recorded, including decision-making time, correctness of diagnostic characterization, and reported confidence. Characteristics observed as the basis of each decision were also solicited. Preferences and comments were reported after the testing, though it is established that preference measures do not always correlate to better performance with a display (Wickens, 1996).

According to information perception theory in cognitive ergonomics, in order for information to be applied it must first be located, then perceived, then integrated with other information into a decision basis. Information as presented in a display must be salient, understandable, and clearly indicative in such a way that it can be compared to patterns in the subject's memory. The main hypothesis was that a display that integrates different data together into a polar star plot would allow the relatively untrained user to more efficiently come up with a more routinely correct diagnosis with less effort than a display of data in separate histogram graphs from which integration must be done as mental work by the user. Having to gather data from different sources and then integrate it can lead to errors in medical decision-making (Gruppen, Wolf, Billi, 1991). Decisions that are made more quickly are usually more accurate in complex problems, so a display that permits the user to find and integrate data more quickly may also contribute to the accuracy of interpretation (Wickens, 1996). There was also a hypothesis that the information the test participants focus on in making their decisions would be the information that delineated the data states in the most salient way.

7.2. Participants

There were 35 Nonexpert participants, ages 21 to 60. Graduate and undergraduate students from the University of Waterloo and The University of Guelph as well as some University of Guelph professors were recruited through email solicitations by department, or sometimes through personal contacts. Represented departments were Systems Design Engineering, Civil Engineering, Human Biology and Nutrition, and Kinesiology. Recruitment did not target any particular gender; 21 participants were female, 14 male. Most of the subjects were familiar with EMG as a subject through their coursework; 10 participants reported they were not familiar with it. Almost all (31) of the participants had studied statistics, though only half (17) reported having studied human physiology.

All participants gave full and informed consent to participate in this study. They were each given $10 as compensation for their time.

7.3. Apparatus and Materials

There was no significant apparatus other than a timer and a clear desk and quiet room. When the beeping of a sports timer seemed so loud as to startle the participants (one participant complained), a JavaScript timer on the computer was adopted in its place.[3] A comparison test was first performed to determine that the comparative times with both timers were approximately the same. The script and training displays, consent form, entrance questionnaire, practice test, task report forms, and exit questionnaire are included in Appendix D. The 18 experimental displays are in Appendix E, along with a key to the code numbers and a list of the display-to-mode assignments and orders for each participant.

Of the 35 participants, the first 19 participated in a form of alpha and beta testing (see Table 7-1) and the last 16 participated in the final testing, which was the basis for the analysis in Chapter 8. The first participant to alpha test the procedure used a different practice test, and provided a lot of feedback on how to improve it. The original practice test was organized differently. The improvements suggested by the pilot tester made it easier to answer the questions while looking at the correct practice displays. Feedback from the first participant also lead to a revision of the training procedure and the introduction of the training displays. Only the final experimental materials (used for the last 16 participants) are included in the appendices (D and E). The first groups of participants gave feedback both through comments and performance on the display designs, which lead to revisions of the polar star and histogram displays. These revisions and the

motivation for them are described in the *Display Designs* section, below. The text display was not modified and remained the same throughout the proceedings.

The groups of participants who viewed each histogram and polar star display design are broken out clearly in Table 7-1 below. Since all of the participants used the same display design for the text mode of display, it is not indicated in the table.

	Polar Star Alpha	Polar Star Beta	Polar Star Final	Histogram Alpha	Histogram Final
Participants 1-6	●			●	
Participants 7-13		●		●	
Participants 14-19		●			●
Participants 20-35			●		●

Table 7-1 Display design allocation across participants. Polar Star Alpha had no normal standard deviation lines. Polar Star Beta had lines indicating the standard deviation of the mean in normal individuals for each characteristic. Polar Star Final was like the Beta except that the size index axis ran from -0.5 to 1.5 instead of 0 to 1. The change from Histogram Alpha to Histogram Final was the addition of lines to indicate means and standard deviations in the distribution of normal individuals. All participants used the same text display.

7.4. Experimental Procedure

An explanatory document and a consent form were presented to each participant, and a questionnaire administered for statistical purposes regarding their experience with computers and EMG. The participant was briefly educated in the domain of the EMG and the six EMG characteristics. This training consisted of a lecture with references to two explicatory diagrams and a table of reference values – normal, myopathic and neuropathic means and standard deviations for the characteristics that would be used in the displays. The complete script and supporting materials used for this process are in Appendix D.

When the training was finished and questions had been answered, participants completed a practice test where they were given three displays to read in each mode (one of each disease type). For each practice display the participant answered a question or two about the display and then characterized it as Normal, Myopathic, Neuropathic, Abnormal, or Unclassifiable (I don't know).

[3] http://www.matcmp.sunynassau.edu/~sherd/Applets/StopWatch/StopWatch.html

Participants were able to refer to the table of reference values during the practice test, though they were warned that they would not be able to refer to it during the experimental tasks. Feedback as to performance on the practice test was given immediately and some users repeated parts of the practice and/or were given clarifying instruction. Participants without statistical training in particular required additional explanation of the histograms. This practice test was done both to train the participant and to collect information about individual skill level in case that turned out to be correlated to the study results.

The experimental task that followed the practice test involved three different modes of display, incorporating five MUAP characteristics and one MUAPT characteristic in each display (described above), which were evaluated through use and interpretation. The participant was shown 18 displays of these characteristics and asked to fill out a task report form (in Appendix D) on how they might characterize the information in each display. Each participant was shown 6 displays in each display mode. The order in which the display modes were presented to the participant was randomized between participants in order to be able to analyze the influence of learning on effectiveness, ease and speed of interpretation. The task report also included questions on confidence and the basis for their decision. The participants were prompted to evaluate their confidence in that answer along a 5-point sliding scale, from "I'm only guessing" to "I'm sure it's correct". The experimenter timed how long it took between when the subject began to examine the display and when they indicated they had reached a decision by saying "done." In some cases the participant didn't pick up on saying "done" so the timing was judged according to when they marked their classification on the task report. In between experimental tasks participants had the option of looking at the Table for Reference.

The participants were also asked to briefly describe what information from the display led to their decision. This was intended to capture indications of which patterns and information were most salient or obvious in each display, and to help us interpret any errors. The participants were allowed to refer to the display while they filled out the task report, except for the last portions of task reports 17 and 18.

For the last two displays presented to the subject additional information was gathered after they were no longer looking at the display. These questions sought to identify more accurately which characteristics the subject had paid attention to and could correctly characterize on an individual basis. This was a memory test and was not included in the first 16 display reports because the researchers did not wish to unduly influence how the participant was using or examining the information presented. After the participant completed the regular task report, he or she was asked to flip the display over and answer the additional questions from memory. Memory

recall can be used as a measure of display effectiveness (Vicente, 1992). Participants were asked to recall both specific values and characterizations (e.g. normal, high, low) for different characteristics in each of the last two task reports.

A final questionnaire solicited comments on the testing and additional comments on the difficulty of interpreting the displays, the subject's preference between the displays, as well as the subject's assessment of their own understanding of the six characteristics.

7.5. Measures and Variables

Independent variables were the data sets, the display modes, and the order of the modes as presented to the participant. Dependent variables between participants and within participants included the speed and accuracy of their decisions, and their reported confidence in each decision. Also comments as to the basis of the participants' decisions were categorized and tallied for each display. Display preferences were another subjective dependent variable.

7.6. Display Designs

The characteristics chosen for the experimental displays were duration, amplitude, the area-amplitude ratio (AAR) or thickness, number of phases, size index, and firing rate per motor unit. Some of them were chosen because they are commonly used in clinical diagnosis (though not always quantitatively) and because most of the necessary reference values of means, standard deviations and correlations for these characteristics were available in the literature. These references were used to design a method for simulating disease and normal data sets for these characteristics (see *Simulating Data*, below).

All participants were shown 8 data sets of each simulated case: Normal data, Neuropathic data, Myopathic data. Two data sets of each case were used for training and practice, and the other 6 were used for the experimental task. These cases (24 in all) were randomly distributed into the 3 display modes: a tabular presentation in plain text, a set of 5 histograms and one scatter graph, and a six-dimensional polar star plot. The randomization process used a Matlab function to put the numbers 1-18 in a random order. Positions 1-6 would be shown in the first mode, 7-12 in the second mode, 13-18 in the last mode. The mode orders were then filtered by hand such that all orders that did not have at least one data set from each state in each mode were thrown out.

All three display modes were designed to show the same information to the greatest extent possible, though each had information the other displays did not. The tables of text (hereafter

called the text display) listed the precise standard deviation, which was not indicated in the other two display modes. The histograms showed the actual distribution of the data, which the other two did not, and (while using the Alpha histogram display) the polar star plot was the only display that indicated where the data mean fell in relation to normal. That comparison is an intrinsic characteristic of how the polar plot is designed.

7.6.1. Text Display

The text display was modeled directly off of the current Muscle Study Results screen in DQEMG. The parameter names, size and font, and the rest of the format followed that screen design. This was based on the rationale that we wanted to test how effective the current design already was for providing information in such a way that it could be integrated and interpreted, and also that we wanted to compare the current information display with the proposed new graphical designs: the histograms and polar plot. The number of digits displayed in numbers in all three displays (two after the decimal point) was based on staying consistent with the original DQEMG Muscle Study Results display. See Figure 7-1 for an example of a text display.

Parameter Type	Mean	Std. Dev
Peak to peak Voltage	324.85	236.89
Duration	8.46	2.71
Phases	4.70	1.78
Area to Amplitude Ratio	1.06	0.43
Size Index	-0.08	0.45
FR Mean/MU	1.95	0.94

Figure 7-1 Text display for myopathic data set.

7.6.2. Histogram Display

The histogram display consisted of five frequency polygons (duration, amplitude, phases, AAR, and size index), similar to the histograms used in the Expert testing but without the actual data points shown on the display. Each bin of the histogram was labeled with a single number indicating the highest range of that bin. The last bin number indicated the outside range of the histogram. Amplitude was graphed on a logarithmic axis so that neuropathic data could be

graphed without changing the axis range; another difference between the Expert test histograms and these histograms was that now the axis range was kept consistent in all displays. The purpose of this change was to avoid introducing another variation in information from one display to another – if the axis were to change with the data a change in axis itself would sometimes be indicative of an abnormal condition. The vertical scale was also kept constant.

There were 8 FR/MU data points in each data set, one for each contraction. That's not enough data points for a histogram, so FR/MU was graphed on a scatter plot along a horizontal axis with random vertical variation introduced to make all the data points visible. The data points were plotted and also the mean was indicated with a vertical line, labeled with the mean value. There were two incarnations of the histogram display, which will be referred to as the alpha histogram and final histogram (or just histogram) displays.

7.6.2.1. Alpha Histogram Display

The alpha histogram display had no reference lines for normal mean and standard deviation. The participants reading this display had to remember normal values as on the text display. The alpha testing of the histogram display involved participants 1 through 13 (see Table 7-1 for the distribution of participants across displays); results of the initial testing with the alpha histogram display were somewhat disappointing, which motivated the redesign to the final histogram display.

During the alpha testing, participants produced an overall error rate of 14% on the histogram display: 4% on normal cases (1 out of 27), 16% on myopathic cases (4/25), and 15% (4/26) on neuropathic cases. This last error rate was a surprise, and was twice the error rate for text displays on neuropathic cases (there were no errors classifying neuropathic cases on the polar star displays by any of the alpha test participants). The alpha histogram displays had been expected to make neuropathic cases clearly visible, especially in the phase and size index histograms, and yet two participants misclassified neuropathic cases as normal and two as myopathic. Only one of those cases was a borderline data set, where 40% of the data in the set was normal (see *Data Simulation*, below).

Two participants complained that the normal information was not shown and one suggested adding a line for the normal mean. One participant commented that the alpha histogram display had an easy layout and another liked how you could look at each variable separately and compare them, but three people categorized it as difficult to read. One of those said there was a lot to look at and another had difficulty characterizing the standard deviation. None of the alpha test participants preferred the alpha histogram display.

A participant who used the alpha histogram display suggested putting normal references on more displays besides the polar star. It was thought that similar graphical indicators of normal means and standard deviations could be added to the histogram display. It occurred to us that the amplitudes were difficult to compare to normal on the logarithmic scale, even though the neuropathic means were all outside two standard deviations from the normal mean. It was thought that if the normal mean and standard deviation were drawn on the histograms for reference that would make it easier to interpret that relation accurately. This idea refers back to the ecological display design guideline that if the user has to infer a relationship between two things it is better to represent that relationship directly in the display. The text display was more problematical; we considered using colors to indicate numbers more than one standard deviation above or below normal, but color would have been a dramatically new visual element, the impact of which might have made it harder to compare the displays. Therefore nothing was changed on the text display.

7.6.2.2. Final Histogram Display

Participants 14 and up were given the final histogram display, which had the normal mean added as a dashed line, slightly shorter than the line for the data set mean, with one standard deviation from the mean indicated to either side of the normal mean with shorter dashed lines. On some normal cases the mean would be right on top of the normal mean, as in the phases histogram below (Figure 7-2).

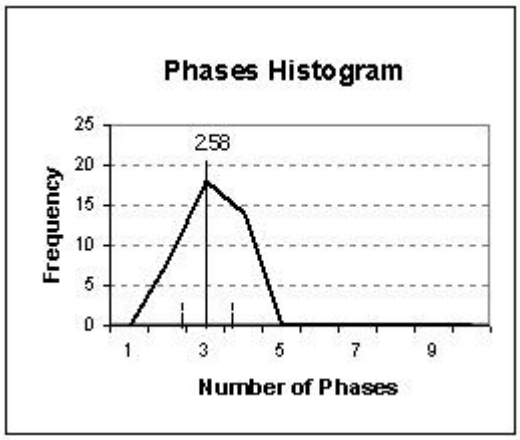
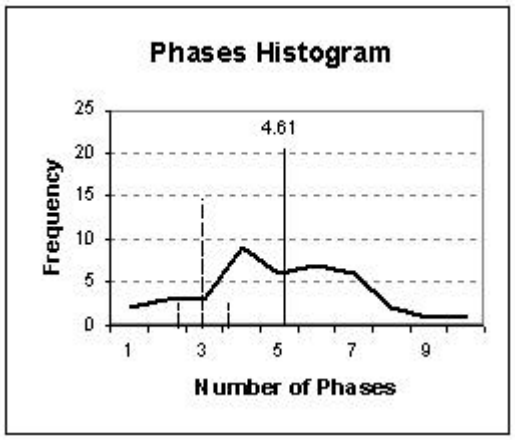


Figure 7-2 Normal phases histogram.

In the abnormal cases the participants were expected to be able to directly perceive whether the data mean was smaller or larger than the normal mean by whether it was to the left or the right of this indicator. The histograms also provided a visual way of estimating and using the standard deviation of the data. Participants were trained to notice that a distribution with a small standard

deviation would result in a tall peak, as in the normal phases distribution, while a large standard deviation would create a wide spread-out appearance, as shown in Figure 7-3.

a. 

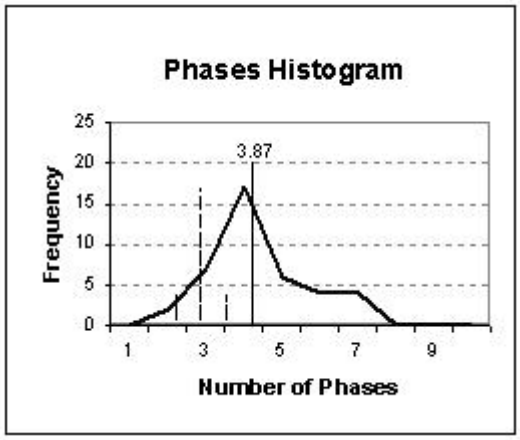
b. 

Figure 7-3 Phases histograms for a) myopathic and b) neuropathic cases. Note that in general, the myopathic standard deviation is larger than the neuropathic standard deviation.

Classification of myopathic cases was expected to be more difficult than classification of neuropathic cases, but both would be assisted by the addition of normal mean and standard deviation reference lines to the histogram display graphs. Since the normal mean and the standard deviation of the mean were at this point already being represented in the polar star displays, this was also done to maintain the comparability of the two displays.

7.6.3. Polar Star Display

A new 6-dimensional polar plot was designed for the Nonexpert user testing. It was possible to create a triangle in the neuropathic case by putting characteristics that were expected to be larger in neuropathy and smaller in myopathy along an axis opposite characteristics that would not change so significantly, and alternating them, so amplitude was on the top vertical with phases on the bottom, size index was on the lower left corner with duration to the upper right, and FR/MU was on the lower right corner with thickness (AAR) to the upper left. Amplitude was displayed on a logarithmic scale. There were four versions of the polar star design, two of which were similar enough to be treated as a single design in this paper.

7.6.3.1. Alpha Polar Star Display

The alpha design plotted the mean of each characteristic on an axis that ranged from 0 to approximately twice the normal mean. Midway from the origin to the outside range, an equilateral polygon was drawn with a dotted line to represent a normal shape. As on the 5-dimensional polar

star, the mean of the data was drawn with a thick line perpendicular to each axis, and a black line was drawn from mean to mean to make a polygon.

The alpha testing of the polar star involved participants 1 through 6. In these participants, the histogram display had the highest incident of erroneous classification (11 %, or 4 out of 36), followed by the text display (8 %: 3 out of 36) and then the polar star display (6%: 2 of 36). It was surprising to the designers that one of the polar star errors was for a normal data set, misclassified as neuropathic – identification of normal in the polar star was expected to be automatic. A possible explanation was that the size index of .76 on that display was large enough to make the display asymmetrical and pull the polygon askew along the size index axis. With the axis ranging from 0 to 1, the standard deviation of the normal mean (approx. 0.25 to 0.75) was very wide for size index. Considering the error identifying a normal case it was decided to indicate the standard deviation from the mean for normal individuals on the next polar star display design.

7.6.3.2. Beta Polar Star Display

The initial beta polar star display (a) indicated the standard deviation of the normal mean with a dotted line drawn parallel to each axis. When the normal standard deviations were introduced to the histogram displays with participant 14, this was changed to a pair of dashed lines perpendicular to the axis (b) in order to be consistent between the polar star and histogram displays.

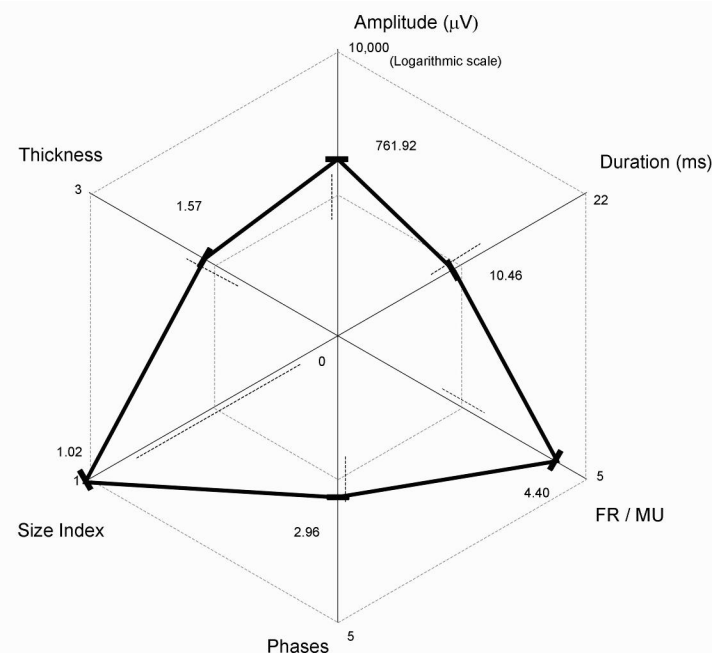


Figure 7-4 A neuropathic case on the beta polar star design (a).

The information included in the display was not changed, so the displays with the two types of standard deviation indicator are all treated as the beta polar star display. The theory was that a visual indication of the normal standard deviation of the mean would prevent errors on normal displays in the future by giving the user a reference with which to recognize that a shape was not sufficiently askew to be abnormal. The decision to do this was also supported by the fact that three participants had suggested adding standard deviation indicators to the polar star. An example of a neuropathic data set drawn on this type of polar star plot is shown in Figure 7-4.

When normal means and standard deviation indicators were added to the histogram display, the standard deviation lines on the polar star display were changed to show the limits with dotted lines that were perpendicular to the axes, in order to be consistent between the two display modes. Participants 14 through 19 used this polar star display design.

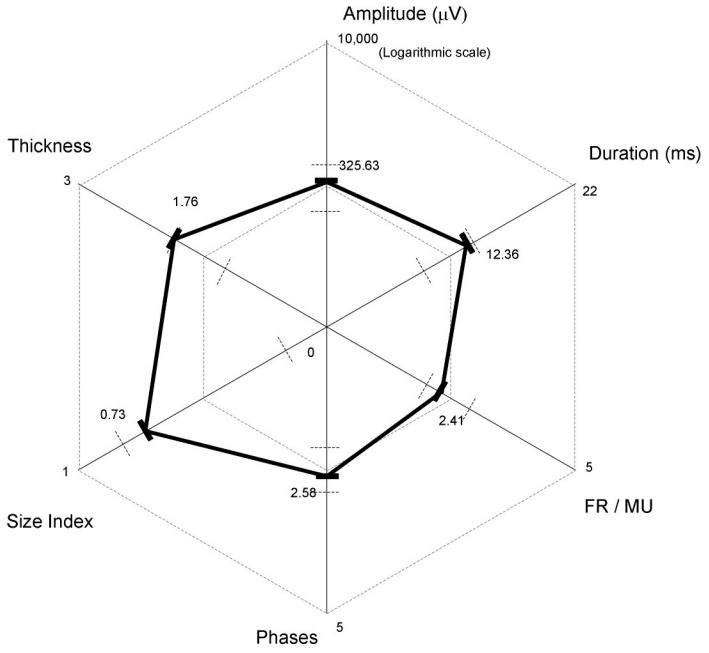


Figure 7-5 A problematic normal case on the revised beta polar star design (b).

Participants continued to misclassify normal cases using the beta polar star design. (The beta polar star display, a and b, was used by participants 7 to 19.) Performances on this display lead to the conclusion that the problem with normal displays had not been resolved. There were 3 misclassifications of normal data on the beta polar star display, for an error rate of 11% (increased from 8% in the alpha test). All of the errors on normal cases on the beta polar star display incorrectly classified them as neuropathic. There were no polar star interpretation errors on myopathic and neuropathic cases in the polar display, but the level of error for normal cases was higher than for text (4%) and histogram (7%) display modes.

At the same time, all but one of the beta testers characterized the polar star display as easy to read and the same group preferred the polar star display over the other two. The one participant who indicated a medium level of difficulty with the polar star (a 2 on the scale from 1 to 4) reported a preference for the text display mode. Reported confidences on the task reports for incorrect polar star classifications were all high, between 3 and 4. Participants who made errors did not comment on any uncertainty regarding those displays or seem to be aware that they might be wrong. With this in mind, it seemed more important than ever to fix the display so that the size index characteristic did not make a normal data set look abnormal.

7.6.3.3. Final Polar Star Display

A new theory was that indicating the standard deviations did not overcome the training that the normal shape should be close to equilateral. The geometric shape being a very directly perceived attribute, the participants might have been making their decisions without considering the standard deviations – participants were judging the polar stars quickly (in an average of 3.9 seconds with a median of 2.3 s) and mainly on shape (see Figure 8-4).

The normal shape would not always be equilateral unless the size index axis range was changed to make the size index standard deviation proportional to the other standard deviations on the graph. An example of a problematic normal case that is skewed rather than symmetrical in the beta design is shown above in Figure 7-5.

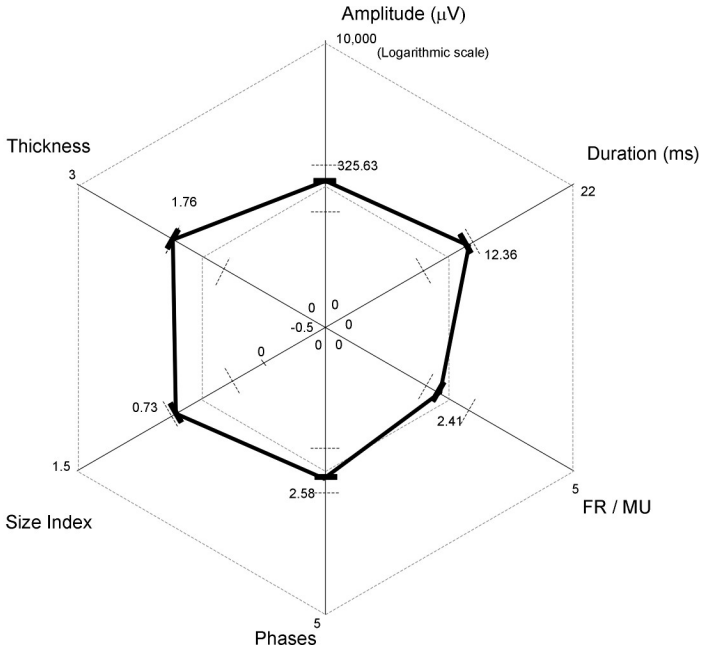


Figure 7-6 Same problematic normal case shown on final polar star design with new size index axis.

Considering that myopathic values of size index could commonly be below 0 and neuropathic values could be up near 2, it was considered reasonable to redesign the size index axis so that the range would be from -0.5 to 1.5. The same problematic normal data set as in figure 4 is shown below on the final polar star design (Figure 7-5). It appears more round and is less extended along the size index axis. This change also meant that the neuropathic displays became slightly less visibly large along the size index axis, and the myopathic displays, which were shaped rather like a kite, were no longer collapsed so far inward along the size index axis. This was not considered a significant drawback. Figure 7-6 shows a myopathic case after the axis change. Before the axis change that data set would have been drawn with size index collapsed inward just past the origin. Participants 20-35 were given the final polar display. An effort was made to recruit a sufficient number of subjects to be able to statistically compare results from before and after this last design change.

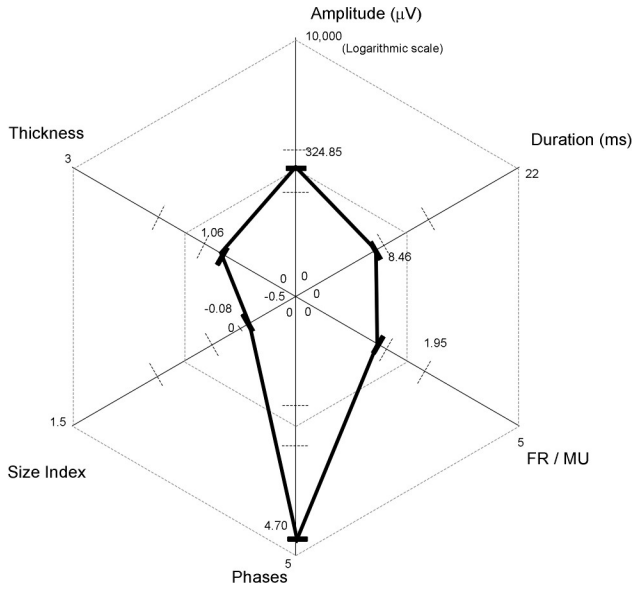


Figure 7-7 A myopathic case in the final polar star display.

7.7. Data Simulation

In order to simulate 24 data sets representing 3 different disease states, a table of reference values was collected for the biceps brachii muscle. Means, standard deviations and correlation coefficients were needed to produce random data that was distributed and correlated like normal, myopathic, and neuropathic data. Some normal reference values that were not available in

published literature were calculated through analysis of data from a 2000 study by Stashuk and Doherty. Some of the results of that study were published (Doherty and Stashuk, 2000).

7.7.1. Normal Correlation Coefficients and Other Normal Biceps Information

In 2000, Dr. Doherty and Dr. Stashuk reported on DQEMG findings in 13 normal cases for the biceps brachii. That report included means, standard deviations, and ranges for mean values for MUAP peak-to-peak amplitude, MUAP duration, MUAP phases, macro amplitude, firing rate, and MUs per contraction. The muscle studies were conducted with subjects aged 23 – 45 who were all healthy with no evidence of neurological disease. Since this data was collected with DQEMG it is especially appropriate to use it as reference data in order to simulate data for user testing.

EMG Data Collection – The EMG signals were sampled at 25 kHz on a Neuroscan A3000 EMG system using an earlier version of the DQEMG application than discussed here. Surface signals were detected with self-adhesive disposable electrodes with a band pass of 5 Hz to 5 kHz. Intramuscular signals were detected with standard concentric needle electrodes and a band pass of 10 Hz to 10 kHz.

Recovering Data – Some of the raw data from the Doherty and Stashuk study was lost when there were problems with a computer and it was sent away to be fixed. A paper printout of the needle (micro) and macro templates was saved, as well as printed versions of the muscle study results summary screens from each study. To complete the record of means and standard deviations for each MU, the micro template information was manually copied into an excel spreadsheet.

Duplication Accuracy – In order to confirm the consistency of the manual copying and to correct typographical errors, the area-to-amplitude ratio was recalculated for each motor unit. This value was then compared to the copied value, identifying most typos in either amplitude or area. In addition, the mean and standard deviation values were calculated for all the MUAP characteristics in each muscle study, and these values were then compared to the muscle study results printouts. This second process identified the most common error, which involved misreading an entire motor unit – in effect, duplicating one and leaving another out. All of these typos were corrected and the resulting data collection is considered to be a reasonably accurate representation of the original data.

Calculations – Since size index and log(amplitude) had not been previously calculated for this data, that calculation was undertaken once the data was input into the spreadsheet.
Size index was calculated as

Size index = 2 (log(amplitude) – log(1000)) + area/amplitude

Means and standard deviations were calculated two ways. First each subject's MUAPs were averaged together and the means of this individual average were calculated, plus the standard deviation of the mean values between the individuals. Second, the standard deviation of each characteristic was calculated for each individual followed by the average of that standard deviation across individuals. Then the whole group of all MUAPs was pooled together and the mean and standard deviation were calculated for each characteristic in the group. The results are in Table 7-2.

	Individual			All MUAPs	
	mean	std dev of the mean	avg. std dev	mean	std dev
Duration (ms)	10.8	1.5	4.25	10.8	4.4
Phases	2.45	0.19	0.77	2.48	0.78
Turns	2.81	0.26	1.19	2.79	1.21
Amplitude (μV)	325.36	83.6	170.5	314.98	182.05
Area	509.27	112.74	283.3	503.78	302.01
AAR	1.63	0.21	0.57	1.65	0.59
Size Index	0.53	0.24	0.67	0.52	0.70
log(amplitude)	2.45	0.10	0.21	2.44	0.23

Table 7-2 Normal biceps MUAP characteristics.

In order to simulate data for normal subjects, the individual means and average standard deviations were considered the most appropriate representative statistic to characterize the distribution within an individual subject. The standard deviation of the mean was used for the polar star display design.

	Duration	Phases	Turns	Amplitude	Area	AAR	Size Index
Phases	0.36						
Turns	0.34	0.54					
Amplitude	0.31	0.24	0.38				
Area	0.70	0.33	0.37	0.75			

AAR	0.71	0.20	0.08	-0.15	0.45		
Size Index	0.82	0.36	0.32	0.49	0.88	0.76	
log(ampl)	0.35	0.30	0.39	0.94	0.77	-0.11	0.56

Table 7-3 Normal biceps MUAP characteristic correlation coefficients.

All 322 MUAP template values were pooled in order to calculate correlation coefficients for duration, phases, turns, amplitude, area, area/amplitude, log(amplitude), and size index (Table 7-3).

7.7.1.1. FR/MU

Quantitative reference values for FR/MU in decomposition-based QEMG were unavailable in the literature. Therefore a study of normal firing rates and firing rates per motor unit was also performed on the data set from Doherty and Stashuk. Since FR/MU is a characteristic of the contraction rather than the MU, calculations had to be done on a contraction-by-contraction basis. First the average firing rate for each contraction was calculated, and then the number of MUs identified as active during each contraction was counted. Taking the first number and dividing by the second, we obtained the following statistics to characterize FR/MU in the biceps brachii of normal subjects. Results of these calculations are collected in Table 7-4.

	Individual			All Contractions	
	mean	std dev of the mean	avg. std dev	mean	std dev
Firing Rate	12.28	1.41	1.04	12.21	2.42
# MUs	5.96	1.95	1.60	5.68	2.59
FR/MU	2.55	1.00	0.85	2.58	1.34

Table 7-4 Normal Biceps Firing Rate, MU count, and Firing Rate per MU (FR/MU) characteristics.

7.7.2. Neurogenic and Myogenic Biceps Information

In two papers Zalewska and Hausmanowa-Petrusewicz (1999, 2000) published means and standard deviations for duration, area, AAR, size index, phases, turns, turns/phases, and irregularity coefficients in the biceps brachii muscle, as well as correlation information. Both studies involved data collected from 20 patients with neurogenic processes and 14 patients with myogenic disorders. Neurogenic and myogenic MUAP characteristic information was taken from these studies. Some information was available in published results and further information was solicited through correspondence with E. Zalewska when necessary.

EMG Data Collection – The EMG signals were sampled at 26.5 kHz by Dantec EMG equipment. Intramuscular signals were detected with concentric needle electrodes with a band pass of 10 Hz to 10 kHz. The resulting signals were transferred to a PC and analyzed using their in-house software.

7.7.2.1. MUAP characteristics

The mean and standard deviation numbers as provided were split into two subcategories in each disease case, according to complexity. Within neurogenic cases, they reported statistics on 97 simple MUAPs and 118 irregular MUAPs and within myogenic cases they reported statistics on 91 simple MUAPs and 150 irregular ones. Since our work intended to consider the distributions of neurogenic and myogenic cases as whole groups, the statistics of these subcategories were combined (see Table 7-5). A weighted average was calculated for the mean:

$$\mu = (\mu_1 * n_1 + \mu_2 * n_2) / (n_1 + n_2)$$

The standard deviations were combined similarly (Harvey, 2001):

$$\sigma_{total} = \text{sqrt}\left(\left(\left(\sigma_1^2 * (n_1 - 1) \right) + \left(\sigma_2^2 * (n_2 - 1) \right) \right) / \left(n_1 + n_2 - 2 \right) \right)$$

	Neuropathic		Myopathic	
	mean	std dev	mean	std dev
Duration	11.527	4.203	7.833	2.804
Phases	3.554	1.432	4.115	1.871
Turns	5.496	2.522	6.262	3.019
Amplitude	1.109	0.859	0.310	0.190
Area	2.08	1.76	0.27	0.144
AAR	1.86	0.77	1.0	0.43
Size Index	1.70	1.03	-0.162	0.469
log(amplitude)	2.921	0.326	2.413	0.261

Table 7-5 Mean and standard deviation values for neuropathic and myopathic MUAP characteristics in the biceps brachii (from Zalewska and Hausmanowa-Petrusewicz, 1999, fig 3 –exact values provided by email).

Correlation coefficients were published in Zalewska and Hausmanowa-Petrusewicz (2000) for all of the MUAP characteristics needed for this study with the exception of log(Amplitude). Data

simulation with amplitude values was attempted but was not successful; amplitude could not be assumed to be Normally distributed. Some of the resultant data points were negative; the simulated values would not have occurred in human subjects. Since the log of amplitude has a more normal distribution, the correlation coefficients of log(amplitude) were obtained from E. Zalewska. The tables below list the correlation coefficients that were used for data simulation for the neuropathic and myopathic cases.

	Amplitude	Duration	Phases	AAR	Size Index
Duration	0.1				
Phases	0.15	0.42			
AAR	0.03	0.76	0.1		
Size Index	0.6	0.65	0.18	0.78	
log(Amplitude)	0.93	0.15	0.18	0.07	0.67

Table 7-6 Neuropathic correlation coefficients for MUAP characteristics in biceps brachii.

	Amplitude	Duration	Phases	AAR	Size Index
Duration	-0.09				
Phases	0.18	0.64			
AAR	-0.49	0.67	0.29		
Size Index	0.62	0.5	0.47	0.32	
log(Amplitude)	0.95	-0.07	0.20	-0.50	0.66

Table 7-7 Myopathic correlation coefficients for MUAP characteristics in biceps brachii.

7.7.2.2. FR/MU

Neuropathic and myopathic FR/MU information was unavailable in published literature. Thanks to ongoing collaboration with London University physicians, the BSDAL had access to a small number of studies from patients who had been diagnosed with myopathic and neuropathic conditions. Of the available muscle studies, the average myopathic FR/MU was 2 and the average neuropathic FR/MU was 6.4. Neuropathic values varied from 5.2 to 13.6.

Concerned that the lab had too small a sample to be sure of the patterns in disease categories, we chose slightly different means and standard deviations to report to the Nonexpert participants and use to simulate the experimental FR/MU data. The objective was to include recruitment information while not making FR/MU too significant a characteristic in decision-making. It was decided that using a neuropathic mean of 6.4 would make the neuropathic FR/MU too distinct from the normal FR/MU, giving it too much weight in the minds of the Nonexpert participants. Similarly a standard deviation such as is suggested by the observed range would be

easily noticeable compared to the normal and myopathic values. For this reason the neuropathic mean and standard deviation used for the Nonexpert user testing were 5.0 +/– 1.6. The myopathic ones were 2 +/– 1.4.

7.7.3. Data Simulation

7.7.3.1. MUAP Characteristics

In order to simulate normal, myopathic and neuropathic data sets, the correlation and standard deviation information had to be combined into a covariance matrix that was subsequently applied in transforming a random set of Normally distributed data into a data set that followed the relevant distribution for the given data class. For each MU data entry, a collection of normally distributed random numbers was generated using the randn function in Matlab. An eigendecomposition of the square root of the appropriate covariance matrix was used to transform into a set of properly correlated normal, myopathic or neuropathic distributions of log(amplitude), duration, phases, AAR and size index. The mean values were then added to offset the simulated distributions to the appropriate mean for each characteristic (Jernigan and Fieguth, 1999). Following this process, the log(amplitude) values were inverted to provide values of amplitude in a lognormal distribution.

A covariance matrix is constructed so that the diagonal elements b_{ii} are standard deviations, squared, of variable i, and the off-diagonal elements each consist of a covariance, . The covariance between two variables is calculated according to the following equation (Jernigan and Fieguth, 1999):

$$b_{ij} = \rho_{ij} * \sigma_1 * \sigma_2$$

Where $\rho_{1,2}$ is the correlation coefficient for variables 1 and 2. The resulting covariance matrix is symmetrical and all values should be real. The log(amplitude), size index correlation coefficient in the neuropathic case had to be decreased by .01 in order for the eigendecomposition to work correctly and not produce imaginary numbers in the covariance matrix. The covariance matrices used for normal, myopathic and neuropathic transformation are included in the data simulation Matlab code files in Appendix C.

Eight normal data sets, eight myopathic and eight neuropathic data sets were simulated. Each data set contained 40 MU data entries. Four of the myopathic data sets and four of the neuropathic sets were designed to simulate “borderline” cases where some of the motor units were as yet unaffected by the disease process. These data sets were composed of 60% diseased MUAPs

and 40% normal ones. Once the data sets were simulated, it was confirmed that the classes of all the data sets could be identified successfully based on a text display of their means and standard deviations. This eliminated any concern that the borderline cases might be too difficult for anyone to classify.

7.7.3.2. FR/MU

Each data set was given 8 FR/MU entries on the supposition that 8 contractions is a reasonable number to expect to produce 40 MUAPTs in a muscle study. Normally the number of contractions required to produce a certain number of MUAPTs will vary, so this is a somewhat artificial assumption. Data points for FR/MU were simulated using the randn function in Matlab, which produces random numbers in a Normal distribution around 0 with a standard deviation of 1. The pattern of deviation in the normal cases from Doherty and Stashuk suggested it would be reasonable to assume a Normal distribution. The random numbers were then multiplied by the appropriate standard deviation. The mean of each case was added to all the data points to shift the data as necessary.

In summary, the following means and standard deviations were used in simulation: normal, 2.6 +/– 0.8; myopathic, 2 +/– 1.4, neuropathic 5 +/– 1.6. For purposes of graphing, some small pseudorandom numbers were generated by hand in order to vary the vertical distribution of the FR/MU data points.

Chapter 8 Nonexpert Testing Results

This chapter reports and examines the results from the Nonexpert user testing on the final text, histogram, and polar star displays. That testing involved participants 20 through 35. Errors have been analyzed by mode and class; time and confidence patterns have been examined between modes. Learning effects and preferences have also been examined. On the practice test the most common errors were in estimating the standard deviation on a histogram and figuring out whether or not a phases histogram indicated polyphasic activity. Participants continued to have difficulty reading the histograms throughout the experiment. The polar star display on the other hand, was considered the easiest to read, and produced the lowest level of error.

The data was checked for correlations with the statistical information from the Entrance questionnaire. The statistical significance of the correlation coefficients were evaluated with a Student's T-test according to a threshold for a 95% confidence level. There were no significant correlations between errors and experience with EMG or computers. There were preference differences between participants according to their program of study in school, but the cause and significance of that is hard to determine. The statistical information collected in the Exit Questionnaires is included in Appendix F.

8.1. Errors

Two of the final 16 participants made no errors at all. Four participants each made one error and five of them made two. Both of the participants who reported they had not studied statistics made at least one error. More than half of the 29 erroneous classifications were made by the four participants who each made three or more errors. The average number of errors per person was 1.8 +/- 1.5 (median 2) overall, with 1 +/- 1.1 (1) errors per person on text mode displays, 0.6 +/- 0.7 (0) on histogram displays, and 0.3 +/- 0.4 (0) for polar stars.

There was no significant correlation of errors with the age or gender of the participant, nor with the number of hours he or she reported using computers on a weekly basis. The significance of correlations was evaluated using the Student's T test for n-2 degrees of freedom. The threshold for statistical significance in the group of 16 participants was a correlation coefficient, ρ, of at least 0.40.

There was no significant correlation between the number of errors a participant made in a whole mode and the position of that mode in the display order. A learning effect therefore seems

unlikely. A table of errors per mode broken out for each participant is included in Appendix F; it includes average times and confidences per mode as well as errors. There was a positive correlation between time taken to decide and the presence of error. Combined with negative correlations between confidence and time on a mode-per-mode basis and between confidence and error for individual tasks, this suggests that participants took more time to decide when they were not confident and the trend was for them to report a lower confidence when they were wrong.

Due to the random allocation of data sets to display modes, not every data set or type of data set was displayed the same number of times in each display mode. Error rates have been calculated based on representation in each mode.

8.1.1. Normal Cases

In the text displays, normal cases were classified correctly 30 out of 31 times, with an error rate of 3%. There was one normal text display inaccurately classified as myopathic. In the histogram displays, normal cases were correctly identified 30 out of 32 times, an error rate of 6%. The erroneous classifications were "abnormal" and "neuropathic". Both erroneous classifications were associated with a high level of reported confidence, so the participants were most likely unaware of their error. To briefly examine the results of all the participants who used the final histogram display design (participants 14 to 35), the finding was that the normal error rate actually increased from 4% on the alpha test to 10% on the final histogram display.

In the polar star displays, normal cases were classified correctly 32 out of 33 times, a 3% error rate. The one error on the final polar star display was unlike the errors on the alpha and beta displays; this one was misclassified as myopathic. The participant reported that amplitude was the sole basis of the decision. The participant who made that error was one of the few participants to report a preference for the text display on the exit questionnaire. This error does not appear to be of the pattern observed in the first two designs, where participants misclassified data sets 5 and 6 as neuropathic; that problem seems to have been eliminated by the size index axis redesign.

The overall error rate on normal cases was 3%, for 4 out of 96 cases misclassified over all three display modes. As shown in Figure 8-1 (next page), two errors in polar star and histogram displays were on classifying the same data set. A different data set was erroneously classified in the text display mode.

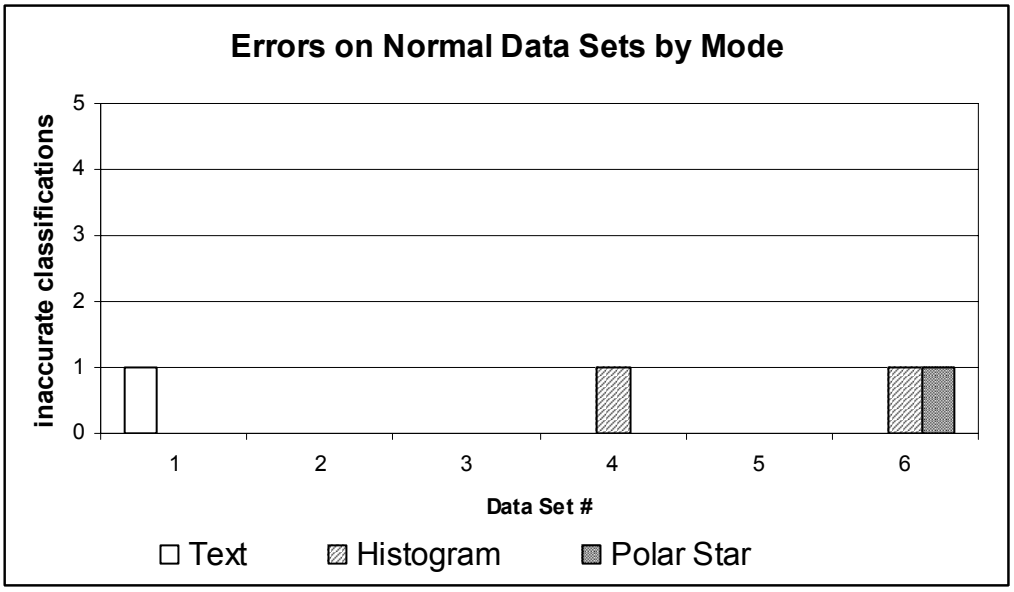


Figure 8-1 Erroneous classifications of normal data sets.

8.1.2. Myopathic Cases

Data sets number 8, 9 and 11 were borderline cases, with 60% of their MUAPs produced according to a myopathic distribution and the rest normal. A graph of the errors in each data set is given in Figure 8-2, organized by display mode.

In the text displays, myopathic cases were classified correctly 25 out of 37 times, which is a 32% error rate. This was the highest error rate of any mode in any class of data. As shown in Figure 8-2, most of the text errors were on the three borderline data sets. Data set number 8 was misclassified three times as normal and once as abnormal, which is at least half right. The person who recognized it as abnormal cited phases as the basis for that classification. Data set 11 was also misclassified three times as normal and once as abnormal (that participant thought the size index might be low for normal), and data set 9 was misclassified three times, as normal. One person who classified set #9 correctly asked if it was borderline. Two participants who misclassified it commented they couldn't decide between normal and myopathic, and one participant had originally classified it correctly, then changed her mind and classified it normal. Data set 7 was also misclassified in the text mode, once, as abnormal, with a reported confidence of 0, or "I'm only guessing."

In the histogram displays, myopathic cases were classified correctly 25 out of 28 times, an 11% error rate. Data set 8 was misclassified once as normal, on the basis of amplitude. Data set

10 was misclassified once, as abnormal, with a low confidence (1) and a long deliberation (65 s). The participant listed amplitude, duration and phases as their basis, overlooking the slightly negative size index. Data set 11 was misclassified once as neuropathic. The participant listed size index as the basis of that decision. Data set 11 was a borderline case, but the mean size index was right at 0, far below the neuropathic mean. This participant made 6 errors overall and may not have had a good memory for numbers and patterns.

In the polar star displays, myopathic cases were classified correctly 29 out of 31 times, for a 6% error rate. One of these errors was a misclassification of data set number 9 as abnormal, based on phases, and the other error misclassified data set number 10 as neuropathic, also based on phases alone. In none of the error cases did the participant refer to the shape of the polar star in their explanation, though they may have used it in classifying other polar stars.

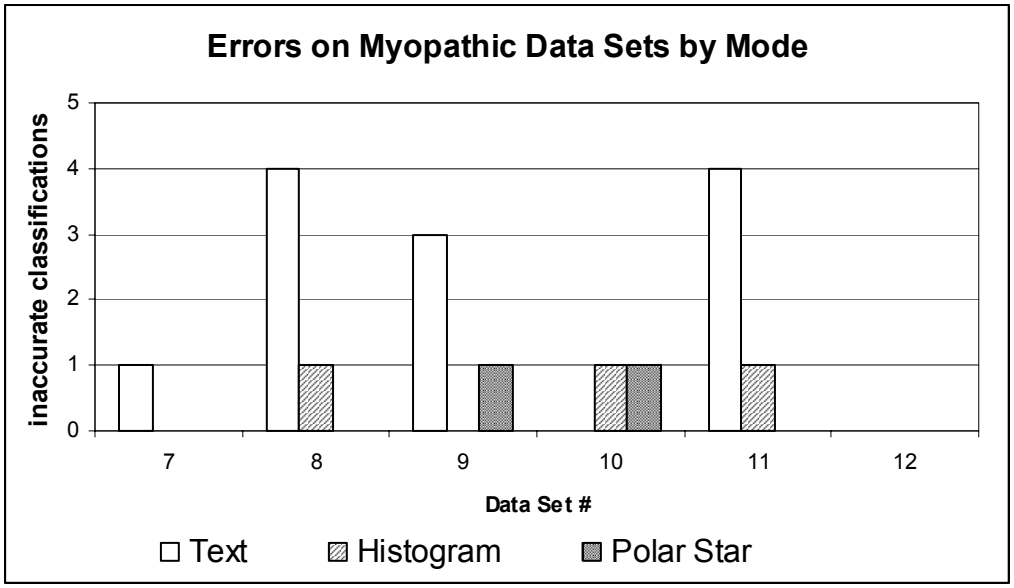


Figure 8-2 Errors on myopathic data sets. Data sets 8, 9, and 11 were borderline.

The total number of errors for myopathic cases in all display modes was 17 out of 96, which is an 18% error rate. Figure 8-2 illustrates that there were more errors on the borderline data sets than on the other three sets of data; 14 of the 17 misclassifications of myopathic cases (82%) were borderline.

8.1.3. Neuropathic Cases

Data sets 13, 16 and 18 were "borderline" data sets, where 60% of their MUAPs were produced according to a neuropathic distribution and the other 40% were normal.

In the text displays, neuropathic data sets were classified correctly 25 out of 28 times, for an error rate of 11%. Data set 13 was misclassified twice, once as normal and once as abnormal. The participant who called it abnormal took longer than average to decide and noted the high amplitude, but still didn't know what it was. The participant who called it normal commented that it "Might be neuropathic." Data set number 18 was misclassified once in text mode, as "unclassifiable," by the same participant who misclassified set 13 as normal and who misclassified 6 of the 18 displays. Four of the 6 misclassifications by this participant were in text mode, which was the first mode for this participant. Previous to this question, the participant had been classifying displays as normal by default if she couldn't tell what they were, explaining when asked that this is what physicians do, isn't it? The participant was asked to label unclassifiable displays as "unclassifiable" from then on and was reminded that the data in each display did actually belong to a specific class.

In the histogram displays, neuropathic cases were classified correctly 32 out of 36 times, an error rate of 11%. Data set 15 was misclassified once as myopathic, though that participant reclassified it correctly after the short term memory questions that followed (this was display number 17 for that participant). Data set number 18 was misclassified three times on the histogram displays, as normal, abnormal, and myopathic. The participant who called it abnormal reported a confidence of only 1.5 and had originally classified it as neuropathic with a confidence of 2.5. The basis reported by that participant was amplitude and FR/MU. The participant who misclassified data set 18 as myopathic only reported looking at phases.

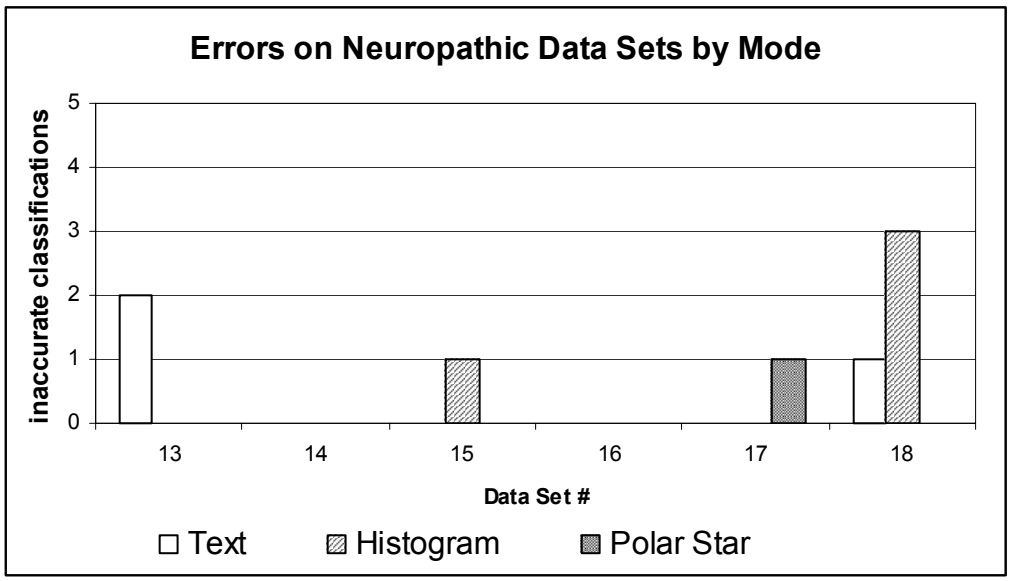


Figure 8-3 Errors on Neuropathic data sets. Sets 13, 16 and 18 were borderline.

In the polar star displays, neuropathic data sets were classified correctly 32 out of 33 times, an error rate of 3%. The single error was a mistake in interpretation rather than perception, as the participant clearly described the shape of the display as a triangle. That participant was running through the polar stars in a fast and offhand manner (judging data set 17 incorrectly as myopathic in just 0.35 seconds) so the tester brought his attention to the fact that he had just classified two very similar displays differently. The participant slowed down and made no further error. The polar star displays do seem to lead to overconfidence in general, however, which will be discussed in more detail below.

The total number of errors for neuropathic cases in all display modes was 8 out of 96, which is an 8% error rate. Again, the borderline cases were misclassified more often – 6 out of 8 errors – as shown in Figure 8-3. Borderline data sets accounted for 20 out of 25 (80%) of the errors on non-normal cases in this study.

	Polar Star			**Histogram**		**Text**
	Alpha	Beta	Final	Alpha	Final	Final
Normal						
% Error	8	11	**3**	4	6	3
Decisions, N	12	27	33	27	32	31
Myopathic						
% Error	9	**0**	6	16	11	32
Decisions, N	11	27	31	25	28	37
Neuropathic						
% Error	**0**	**0**	3	15	11	11
Decisions, N	13	24	32	26	36	28
Total:	5.7	3.7	4	12	9	17

Table 8-1 Error Rates (%) for categorization of displays in each mode of Nonexpert testing. In the polar star mode, the final design introduced a different axis for size index. For the histogram the final design introduced reference lines for normal mean and standard deviation. The text display did not change. The lowest error rate in each disease state is marked in bold.

As shown in Table 8-1, the text mode of display had an overall error rate of 17%. Participants had the hardest time identifying myopathic cases using the text display. Omitting the beta polar star design, the myopathic case caused the most difficulty in characterizing all modes of display, but the text display produced the worst performance.

The error rate for the alpha histogram display was 12 % while the final histogram display error rate was only 9% in the last 16 participants, which appears to be a slight improvement.

However, the added complexity of the histogram design with normal reference lines made histogram performance worse on normal cases, and participants tended to complain about it more while characterizing it as being more difficult. It is inconclusive what the best histogram design was.

On the polar star displays the overall error rate was 4% on the final design. The final size index axis design successfully reduced the error on normal cases from 11% on the beta (4) to 3% (1). The difference in performance between the beta and the final polar star designs in myopathic cases may indicate that the reduced salience of the new myopathic shape was more significant than expected. The change in errors on neuropathic cases amounts to only one more error and is not considered a significant variation, so overall the performance on the final design is considered superior to performance on the alpha and beta polar star designs.

8.2. Analysis of variance

The number of errors each user made in each mode were summated, and the average decision times and confidence in each mode were calculated (this data is tabulated in Appendix F). Because multiple data points represented the performance of the same person, individual variance within subjects had to be accounted for before the variance between mode groups could be identified. A repeated measures analysis of variance (ANOVA) was performed on the per-participant error sums and average decision times and reported confidences per mode. The residuals were plotted and checked for trends during this process to ensure that assumptions for the adequacy of the model were correct. The ANOVA analysis checks the null hypothesis, which is that the means for the different mode groups are the same or not significantly different. This data set has three data points per participant; one for each mode. That makes the whole number of data points 48. The number of different groups is 3: text, histogram, and polar star. The number of elements per group is 16. The model for the ANOVA therefore used 15 degrees of freedom for participants, 2 degrees of freedom for groups, and 30 degrees of freedom (d.f.) for errors, which is the corrected total number of data points, 47, minus the d.f. of the rest of the model, 17 (Box, Hunter and Hunter, 1978).

An SAS system was used to carry out a repeated measures ANOVA with LSD posthoc analysis for errors, time and confidence. A $P<0.05$ level of significance was used as a threshold. The average number of errors per user in text mode was 1, the average in histogram mode was 0.56, and the average in polar star mode was 0.25. The posthoc analysis for errors grouped the three modes into two pairs of means that were not statistically different. The mean error in text

mode and the mean error in histogram mode were not significantly different from one another, and the mean error in histogram mode and the mean error in polar star mode did not show a statistically significant difference. However, the hypothesis that text mode error and polar star mode error were statistically the same was demonstrated to be false (F=3.9, P=0.0312). The polar star displays produced a significantly smaller average number of errors per user than the text mode displays. The natural log of the average decision time per mode was used to normalize the time variable. A significant difference in the mean decision time was found between the histogram display mode and the other two modes (F=4.27; P=0.0234). The average decision time for the histogram mode was therefore significantly longer than for the other display modes. There was no statistically significant difference in average reported confidences from one display mode to another.

8.3. Decision Times

Decision time for characterizing displays averaged just 10.9 seconds, ranging from 0.05 s to 1 min, 57.3 s. The polar star displays were interpreted in the least amount of time.

The text displays took participants an average of 13.8 seconds to classify, with a median time of 10.3 seconds. Minimum decision making time in text mode was 1.3 seconds, while the maximum was 117.3 seconds, or 1 minute, 57.3 seconds. The histogram displays were interpreted, on average, in 12.3 seconds (median 10.5 s). Minimum and maximum decision times in histogram display mode were 1.4 seconds and 65.9 seconds (1 min, 5.9 s), respectively. Polar star displays took an average of 6.6 seconds to classify, with a median below that of 2.9 seconds. Some users regularly took less than a second to make their decision on the polar star displays. The decision time range was 0.05 s - 60.3 s (1 min, 0.3 s).

The ANOVA above confirms that the histogram average time was significantly longer than for the other two modes. The times were adjusted with a natural log transformation for the ANOVA.

8.4. Confidence

The trend on confidence overall was for confidence to be lower in cases of error. However, the correlation between error and confidence on polar star displays alone was an insignificant 0.03. None of the 4 errors on polar star displays were associated with a reported confidence lower than 3

(while other correct polar star classifications had confidences as low as 0.6 and 1.5). This overconfidence on polar stars may be an issue of concern.

8.5. Salient Characteristics

One hypothesis in this study was that the participants would naturally pay attention to the signal characteristics that were more salient or helped them most in distinguishing between classes of data. During the training an effort was made to take an equal amount of time and mention the normal values for each characteristic. It is possible that the training strongly affected which characteristics participants chose to pay attention to, but their reported attention still seems worth examining. A comparison of cited characteristics organized by mode is in Figure 8-4. Responses from all 35 participants have been pooled to show trends more clearly.

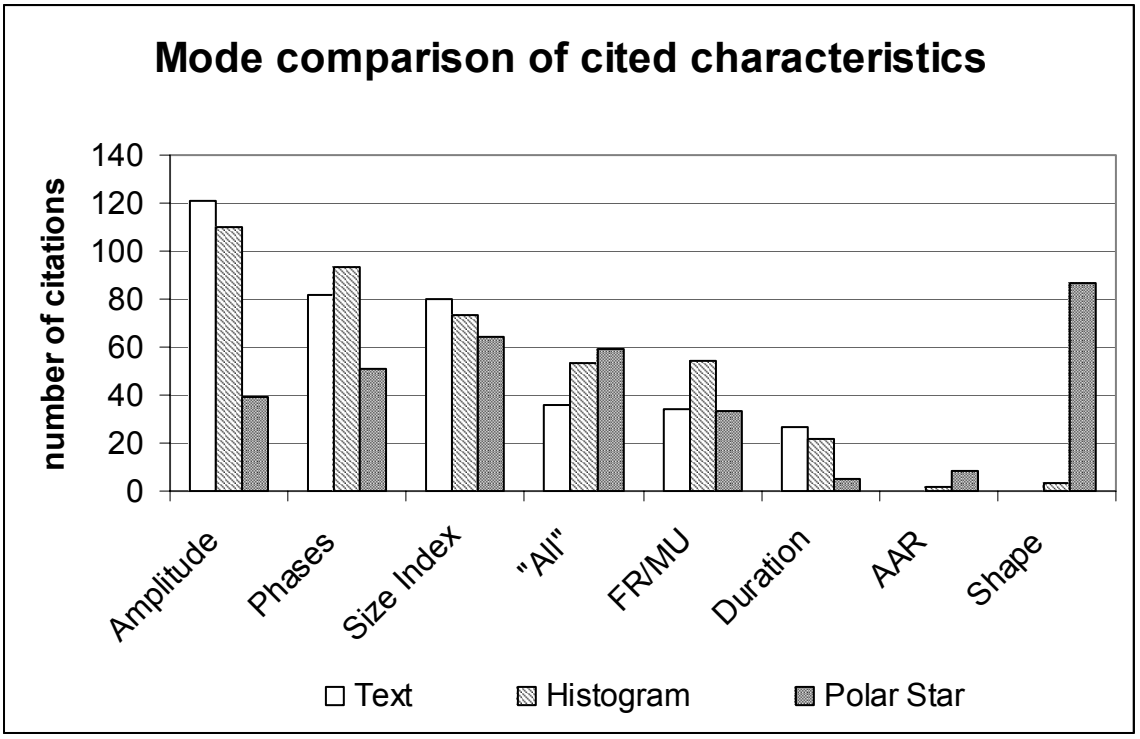


Figure 8-4 Characteristics cited by participants as the basis of their display characterization, according to display mode.

The characteristics that were most commonly identified when the Nonexpert participants reported the basis of their decisions were amplitude (270 times), phases (226), and size index (217) times. On 148 task reports the participants made a general statement like "They all look normal," sometimes in conjunction with mentioning specific characteristics. FR/MU was referenced often (121 times), followed by duration (54) and thickness or AAR (10). Which

characteristics were mentioned also varied by mode (see chart). Participants listed shape in 87 out of 210 decisions on polar stars. Those who did not list shape on polar stars most likely listed size index (54), phases (40) or amplitude (31), often together. The large normal standard deviation of duration and the high overlap in duration values of normal, neuropathic and myopathic distributions could be responsible for the low use of duration for decision-making. AAR also suffers from close distributions. It might be possible to improve the separability of patterns on an AAR graph by making the origin of the axis something larger than zero.

8.6. User Preferences and Difficulty Comments

User preferences were clear in this study. Three out of the 16 participants using the final displays preferred the text display to the other two, and 13 participants preferred the polar star. No participants preferred the histograms. The polar star plot was a clear favorite among participants, with 81% of participants stating a preference for it.

Those who preferred the text displays commented that they were standard and familiar and all of them designated that display mode easy to read. One of them complemented the large font. And one said the "relative variability" was reported most clearly in that display. Of those three participants who preferred text over the other two modes, one made two errors in text mode and one reading a polar star, one only made errors (2) in text mode, and one made 1 error on a histogram display. This demonstrates the well-known usability testing adage that preference does not always align with performance. It may be worth noting that none of the three participants who preferred the text mode were in engineering; the engineering students who participated in the study all preferred the polar star plot. It may be that engineering students are more familiar with graphical displays of technical information. Out of the sixteen final participants, 6 called the text mode "easy" to read, a 1 on the 4-point scale, while 5 called it medium (2), 4 called it difficult (3), and one called it challenging (4). Eight participants (out of all 35) commented that they had to rely on memory to compare the text table to normal values, and that numbers were hard to remember.

Though some participants may prefer the text display to the other two display modes, a display design for medical purposes must place performance before preference. However, there are valuable design inputs to be taken from the responses of those who preferred the text displays. Chief among them are that the font should be large and legible and the standard deviation should be explicitly reported somewhere in a QEMG information display.

Regarding the histogram display, four of the participants said that it had too much stuff on it to be easily read, and three participants reported that they were not familiar with using histograms, which lowered their confidence and made it more difficult. One participant complained that the bin numbers were not very obvious. Only 1 of the final participants characterized the final histogram display as easy to read; 8 said it was medium, 6 called it difficult, and only one called it challenging, for an average reported difficulty of 2.5. Two participants complemented how the histogram display made it possible to compare the standard deviation and the mean. Interestingly enough, in the alpha test for the histograms the average difficulty rating of the histograms was only 1.9 while for the final histogram display the average reported difficulty (across participants 14 to 35) went up to 2.4. At the same time, the average reported difficulty of the text displays went down from 2.3 to 1.9. This change may reflect a sense of comparison on the part of the participants. All but 6 out of 22 (28%) of the participants using the final histogram display rated it as hard or harder to use than the text display, while 5 out of 13 (39%) of the alpha testers thought the text display was more difficult.

The polar star plot was designated "easy" to read by 13 out of the final 16 participants. Two participants called it "medium," and one called it challenging, commenting "I have never seen polar star graphs before". That last participant was the only one of the participants who made an error using the polar star display (4 participants in the final session, 8 participants out of all 35) to classify it as anything but easy, which confirms the impression from reported confidences that most of those who made an error using the polar stars were not aware that they did so.

Another valuable comment was that for most of the characteristics 2 significant digits following the decimal point is too many significant digits.

Chapter 9 Discussion

9.1. Acceptability

In the Introduction, the merits of this research were discussed in terms of the potential for successfully introducing a useful application into the medical field. In order for a new application or technology to be accepted, it needs to achieve a basic level of success in a combination of areas that lead to acceptance. Among other things, the Expert user testing made it clear that the DQEMG application has a ways to go on a number of these fronts.

Acceptability can be modeled as a combination of social acceptance and practical usability, as in Figure 9-1 (Neilsen 1993). Practical usability can then be broken down into reliability, compatibility, cost, usefulness and other attributes, where usefulness is again divided into usability and utility. The "Expert" user testing focused on usability, while the Nonexpert display testing explored some aspects of utility. Therefore, the acceptability model seems like an appropriate framework in which to discuss the performance results, participant comments, and other observations gathered in the course of this research.

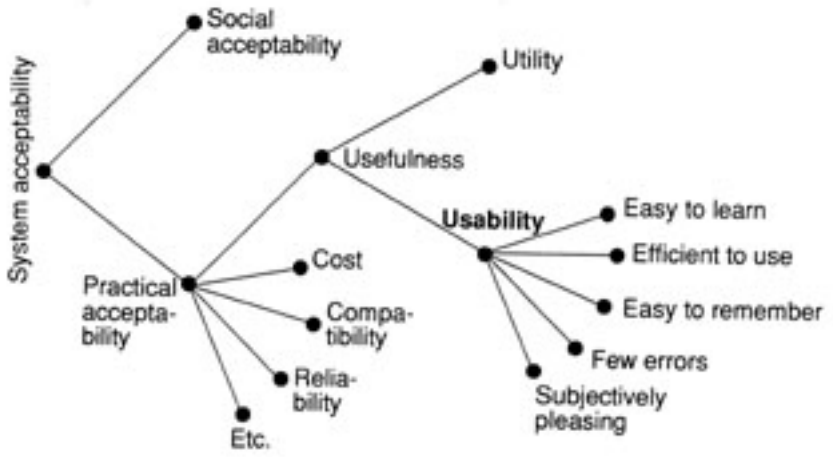


Figure 9-1 A Model of Acceptability (reproduced from Neilsen, 1993)

9.1.1. Social acceptance

There is increasing concern in the EMG community over the use of technicians to do EMG studies. Conventional wisdom is that only a skilled physician can administer and interpret such

studies, and a mere technician should not do them (Preston and Shapiro, 1998). To the extent that DQEMG might make it easier to collect, save, analyze and interpret EMG it might support the efforts of technicians to take a more active role, and for hospitals to save money by letting them. For this application to be socially acceptable we should probably aim to avoid that association.

At the same time, the Nonexpert user testing results support the idea that inexperienced users can consistently interpret the polar star plots with high accuracy. This conclusion is tempered, however, by the understanding that the data for the Nonexpert testing was simulated and therefore did not suffer from some of the variation by age, gender, strength of muscle contraction, etc. that can affect the correct interpretation of a muscle study. Furthermore, the designation of a study into a class of neuromuscular disease is obviously not as complex a task as considering all the possible neuromuscular diseases the patient might be suffering from. A trained physician is needed to make such detailed analysis.

9.1.2. Practical Acceptability

9.1.2.1. Cost

We had two out of six expert subjects inquire about the cost of the application. Since it is still in a research stage, the cost is unknown. A limited version of the application that decomposes and analyzes data but does not allow the user to save and retrieve studies is currently bundled with the Comperio system. The eventual cost of the DQEMG program is expected to be reasonable, so this will probably not be a barrier to acceptance.

9.1.2.2. Compatibility

The DQEMG application only runs on a PC and has better performance if the PC is equipped with at least a Pentium IV processor. DQEMG does not offer any data export capabilities, and it does not support the windows utilities of selecting and copying text from the display, so it is not compatible with any other record-keeping software. However, since some of the file types DQEMG uses to store data (the .prm files and study.txt) are simply tab-delimited text files, any application that can import ASCII text can read the data files. An export function could be created, or a set of headers added to the .prm file, in order to facilitate compatibility.

9.1.2.3. Reliability

There are two factors of reliability that can be considered for this application. The first has to do with whether or not it reliably runs. In the course of the expert testing, users running the DQEMG program suffered from a number of dysfunctions where data was lost and either the application or the computer had to be restarted and the user had to repeat the task or the study. Over half of our expert test sessions were interrupted by at least one instance where the application or the computer crashed. Half of the expert users qualified their interest in using the program professionally by raising questions about its reliability. This included one of the two users who did not suffer a system crash during the test. Some of the problems have since been eliminated, but a functional testing protocol will be able to demonstrate that the program works reliably.

The second type of reliability has to do with how consistently the decomposition algorithm works to achieve the expected goal. This issue has been addressed elsewhere (Stashuk, 1999, Doherty and Stashuk, 2000). The DQEMG application is consistent in identifying motor unit trains that are clearly separable and follow certain thresholds of amplitude and number of firing times. It may exclude action potentials that are superimposed on others as well as motor units with low firing rates. The Biological Signal Detection and Analysis group is researching a method for resolving superimpositions, but the algorithm could yet be improved. In any case, the program does not need to identify every single MUAP in order to provide a highly valuable set of representational information to the physician.

The lack of reliability with which DQEMG identifies the onset and offset landmarks in MUAP templates causes extra work in marker editing. If that reliability could be improved, the application would be substantially more acceptable.

9.1.2.4. Usefulness

Utility

As one user put it, he would want clear evidence that using DQEMG would change diagnosis, treatment, and outcome; effective use of the DQEMG program is expected to reduce error in diagnosis and also to make the determination of treatment easier and more effective, but these advantages are as yet unproven. We are aware that many of the physicians consider its acceptability in terms of time to use versus improvement in diagnostic yield. Our expert commentators seemed hopeful but unsure on this count. Indications are that the application would need to include information on quantitative norms in order to have the greatest utility, since

physicians do not store the necessary quantitative references in their heads, and their mostly qualitative current practice and training does not prepare them to interpret quantitative information. Since it does not include this normative information for each applicable muscle right now, we must conclude that the current utility of the application is probably pretty low compared to what it could be. A third of our "Expert" users characterized a muscle study as normal when according to the numbers it showed clear signs of chronic reinnervation, with a mean amplitude twice the outside range of normal, some amplitudes four times the outside range of normal, and a mean size index above 3.

The graphical displays explored in the Nonexpert testing that give normative references as an inherent part of their design would have a high utility in this regard. Especially the polar plot, with its demonstrated high effectiveness and low error rate. There were no abnormal cases incorrectly classified as normal in the Nonexpert testing using the final polar star display; an abnormal condition was thus demonstrated to be clearly distinguishable from normal.

Usability

Easy to Learn

We received mixed reviews on this count. It took less than an hour for the coached subjects in the Expert testing to approach a reasonable level of proficiency with DQEMG. This rate of learning seemed acceptable to the participants, who were a sample of the target population. Since a coach will not be available in all cases, however, the application as it stands is lacking in this aspect. The documentation that was written for the research study will be incorporated into the interface, and more would be better.

Improved clarity in labels and workflow are also necessary to help the novice get started. It was believed that mouseover instructions on buttons would help the novice to get oriented, but we did not observe any users spending time investigating or making use of such messages for navigational cues during the testing.

Efficient to use

The DQEMG program is not really efficient to use, especially not on the time scale our target population is looking for. Marker editing alone took the Expert participants between 8 and 24 minutes. At least half those participants expressed concern and/or frustration with how much time it took.

However, collecting the muscle study is an efficient process (when the system does not crash) and the graphical information display presents a summary of the Muscle Study in a highly efficient manner. One Expert testing participant commented once he saw the graphical display that he would probably just go straight to it if he were regularly using the program. The Nonexpert testing demonstrated that polar plots can be interpreted more quickly than the current text display, which is in keeping with the theory of ecological perception that underlies their design. If the system were made more reliable and the time required for marker editing could be reduced dramatically, the high efficiency of the information display would make extracting diagnostic information from quantitative EMG a more efficient process.

Few Errors

Users consistently made errors in finding and using the "Edit" and "Markers" buttons for the first and second times, and in changing the scale of the display. They often seemed to have trouble getting back to the screen where they had been before, or on to the next screen they need to see. Some of these things smooth out with practice, but there are mistakes that are not eliminated with experience, possibly indicating that procedures are non-intuitive and/or difficult to remember. There is a more detailed discussion of suggestions for fixing this problem in the *design suggestions* section that follows.

As has already been discussed, there were a significant number of errors interpreting the archived data in the Expert user testing. The Nonexpert testing was designed to help determine how to improve the information display to eliminate some or all of those types of error. The six-dimensional polar star display produced a significantly smaller error rate in Nonexpert participants than the text display (which was designed following the current Muscle Study Results display). The overall error rate using the polar star displays was also lower than the existing error rate of physicians interpreting EMG in the traditional qualitative way (Johnson, et al, 1976).

Easy to set up and maintain

The release version of DQEMG comes bundled with the Comperio system and requires no particular setup. The computer display has to be set to one particular screen resolution (1280 x 1024) for everything to appear and function the way it is designed to function, but that only needs to be done once. In the future there will be maintenance issues for the database and compare functions, but those are topics for further research.

Subjectively Pleasing

The DQEMG application is generally pleasing, and a number of Expert participants complemented it as user friendly or said that they liked the graphics. Nonexpert participants also complemented the polar star design. However, there were also complaints from the Expert Participants that one of the colors used in the application (a gray color) is too dim and hard too see against the black background, and that the display and interface can be too complex or crowded. There were many comments from the Expert participants that indicated that they were confused, and they were of course not pleased to be confused. There was information in the results and summary displays that most users said they would never use, and the program may be able to be simplified in other ways. Again, please refer to the *design suggestions* section.

9.2. Limitations

9.2.1. Identifying Errors in Expert Testing

What is an error in MUAP landmark placement? Originally we planned to investigate the question of how many of our participants made errors in marker placement, but it turned out to be more subjective than expected; there were many instances where there was no clear-cut, logical reason for thinking one marker position was more correct than another. A couple of the participants even complained about how subjective it was. This question became slightly irrelevant when a combination of crashing computers and the one computer being stolen lost the majority of the data representing Expert User changes to the archived study and made it impossible for us to do a detailed analysis of MUAP landmark placement. Still, an inability to state clearly when a participant made an editing action that was erroneous or unhelpful limits the conclusions we might draw. Furthermore, there was some uncertainty as to whether or not it was an error when a participant failed to locate the most efficient path to do what they wanted to do next, but found another way to get there. Issues like these limited the clarity with which we were able to categorize events and comments in the Expert testing.

9.2.2. Using Simulated Data

A major limitation of the Nonexpert testing was the nature of the simulated data. The MUAP characteristics were simulated based on assumptions of Gaussian distributions and without specifically knowing whether the contraction levels of the normal and diseased studies on which the simulation was based were actually comparable. More dramatically, the means and standard

deviations of neuropathic and myopathic distributions of FR/MU were designated based on the known patterns of only a few cases. Though the estimated mean and possible variation of neuropathic FRs/MU were reduced in an effort to limit the influence of that parameter and avoid skewing the results of the Nonexpert research project, a few characterization decisions were reported to be made solely on the basis of FR/MU and the pattern of that parameter contributed in a very definite way to the distinctive neuropathic shape of the polar star graph. We can only rest on our confidence that FR/MU or some other quantitative representation of recruitment would discriminate as clearly between normal and a myopathic or neuropathic process. Similarly, it is not possible to accurately assess the usefulness of the histogram displays in a scenario where the data has no meaningful outliers. The model of a Normal distribution in the data simulation process did not increase the presence of outliers in abnormal conditions, such as Stålberg, Bischoff and Falck suggest is the case in real life (1998). In situations with real data the histogram displays might provide an advantage they did not produce in this experiment.

9.2.3. Variations in Display Design

Another major limitation of the Nonexpert testing is the fact that the text mode of display had no indicators that gave a normal reference, while the final histogram and polar star displays both had indications of normal means and standard deviations so that the data set shown in the display could be compared to them. However, this makes it even more impressive that the text display caused fewer errors than the histograms in all but the myopathic cases, and was preferred by a group of participants. Some other variables such as font size for data and range labels probably should have been kept more consistent as well. The mean value was probably easiest to read on the text display because the histogram and polar star displays both used a font four points smaller to label the mean. This was done to make the labels fit better into the graphs, but probably it was more important to be consistent. One tends to conclude that a polar star design with larger data labels would be even more effective than the final design in this study, so this limitation does not affect our conclusions that the polar star display is the most effective display. There is more room on the polar star plot to increase text sizes than on the Histogram display, but these obsdervations still lead to the expectation that histogram display effectiveness would be improved by the use of larger labels for mean values.

9.3. Future Research

9.3.1. How Necessary is Marker Editing?

One avenue of investigation that merits attention is the question of how necessary the human landmark editing really is. It has not been established whether or not the difference in parameter values and possible conclusions from the application information are significantly altered by this tedious and time-consuming process. It seems likely that they would be, but in the interest of science a small study on that question ought to be conducted. At least some sort of recommendations on how small a change is worth making could be developed and included in the user documentation in order to reduce the tendency to micro edit landmark positions that were observed during the Expert testing. Since the time required for marker editing is something that experienced DQEMG users consistently complained about in the Expert study, determining what part of that time spent has an effect at a level that might change diagnostic yield or other conclusions is a study that ought to be done.

9.3.2. MUAP Landmark Algorithm

The Expert testing participants found and commented that the landmarks that were most commonly inaccurately placed were the offset or end marker and, less often, the onset marker. The onset and offset identification algorithms could be changed to filter the ends of the MUAP signal before assigning the end-point, so it would locate the end of an estimated curve from the first or last peak to the baseline instead of getting caught up in noise and variation that obscures the overall shape of that curve. Reducing the necessity of marker editing by increasing the reliability of onset and offset identification would make a significant improvement in the acceptability and usability of the DQEMG application.

9.3.3. Recruitment and Firing Rate

Comments by Expert users indicate that Firing Rate or even FR/MU might still not be the ideal measure of recruitment. Reporting the mean firing rate in particular was questioned. DQEMG could potentially report the firing rates of active motor units at the time of recruitment of the next motor unit, as well as how many motor units were active when each further motor unit was recruited. Investigation into other quantitative measures for recruitment that might be able to identify both early and reduced recruitment in a satisfactory manner is ongoing.

9.3.4. Multiple Muscle Study Comparison

For neurological studies clinicians may wish to compare multiple muscle points along the same nerve. The DQEMG application could provide a display that does this, for instance a concentric polar star display with muscles along one nerve branch characterized from the spine outward. This suggestion was greeted with some enthusiasm by Dr. Doherty, but the DQEMG application is not currently structured so as to facilitate such a display.

Another kind of muscle study comparison is a longitudinal comparison on the same muscle in the same patient at different points in time. As described in chapter 2, currently physicians compare quantitative reports over time but do not retain the actual waveforms for future comparison. DQEMG samples the waveform and could provide displays that show a dated series of studies side-by-side. A graphical display of this series might be especially effective; the Expert testing and the Nonexpert testing both indicate that graphical displays are both popular and effective. The especially important information in a longitudinal comparison is the relationship of one data set to another. Trends and relationships could be represented directly in a graphical display, as the relationship to normal is shown in the polar star.

9.3.5. Use of Color Coding, Normal References

If a thorough survey of quantitative reference values for different muscles were to have been conducted, the DQEMG application could use that knowledge base to compare collected data to normal values and/or provide indications of disease category ranges. Color is one of the visual indicators that could be used to add this information to the display without necessarily adding whole new display elements. The text-based display could be augmented by the use of color to highlight values that were beyond a couple of standard deviations away from the normal mean. Color could also be incorporated in a more meaningful way into graphical displays such as histograms and polar plots, by coloring the mean indicators or the areas under certain parts of the graphs. The most effective way in which to do this would require some research. One of the Nonexpert users suggested using a color scale of yellow through orange to red to symbolize how far off the muscle study values are from accepted normal values. Red could indicate that a number is more than twice the standard deviation from normal while orange indicates something between one and two standard deviations above normal. Two other colors could be chosen to indicate values one and two standard deviations below the normal mean.

9.3.6. Collecting and reporting Qualitative Assessment

During the original system study we observed that qualitative assessments of insertional activity and EMG recruitment and activity are currently being recorded into a report the physician makes to the referring physician, that gets filed on paper separately from the EMG data. The EMG/EP application on the Comperio system provides an interface so the user can input qualitative assessments of many characteristics of the signal while viewing the raw EMG, but the input scales are poorly designed and so far users are not inclined to use them. If these characterizations were imported into the DQEMG application, they could be saved and reported alongside the quantitative analysis. Perhaps a joint project could evaluate the input scales and improve the interface by changing this to a standard used in clinical labs. With standardized input scales, reports on qualitative assessments of things DQEMG does not capture, such as insertional or spontaneous activity and positive sharp waves could be incorporated into the DQEMG report. This would enhance the completeness of this report, especially in consideration of a longitudinal comparison of one muscle study to a later study on the same patient.

9.4. Design Suggestions

9.4.1. User Documentation

Eventually there should be user documentation on functions as well as developer documentation on classes. Documentation and assistance for users should include an explanation of how to get started using DQEMG (Getting Started with DQEMG, in Appendix A-4) and more detailed help on the functionality and how to use the interface components. In particular, the three major steps of evaluating the decomposition, editing the landmarks and viewing the results should each get a heading in a help index. The section on editing the landmarks should explain the expected relationship between macro and micro MUAPs; for example, the user might use the negative peak onset of the macro to help judge where the micro onset ought to be located, or vice versa. The purpose of re-decomposing a contraction with different settings should also be explained, since this would not be something the typical physician would be familiar with. Since this function is mostly useful for research purposes rather than clinical use, it might also be good to relocate it to a menu rather than the main toolbar.

9.4.2. Labels, Colors and Keys

In all screens that use it, the Edit button should be renamed "Exclude" with a mouseover of "exclude graph data from the study." Screens, graphs and micro MUAP statistics should be consistently labeled Needle or Micro but not both (one or the other) as they are now. There more consistently be a key on the screen to explain the meanings of colors and data. The Decomposition screen should have a color key that indicates the meaning of the colors yellow and gray. Furthermore, the column of MUAPT numbers on the left should be given the header "MUP TRAIN #". On the Raster and Details screens, the numbers should be labeled, with units given.

Throughout the application, the color gray should be changed to a color that contrasts more strongly against black. On the markers editing screen, it should be made clear that the macro or micro landmark that cannot be changed from that screen is just indicated and can't be moved – it should look less like a marker. One participant suggested it be made into a vertical line the entire height of the MUAP template display area. Green might be an appropriate color for that line, since it has been used elsewhere to indicate landmarks such as mean values.

It was unclear how aware the users were of the scale of the Marker editing display and how it related to the standard they are accustomed to in such a way that they could qualitatively assess MUAPs while editing the markers. When normative values are available for a particular muscle, it may be valuable to add a horizontal line to the template graph to indicate the normal average amplitude, or otherwise label the graph so the user can make a qualitative comparison to normal regardless of the scale the display is set on. This could improve the quality of the user's information gathering as they interact with the data.

9.4.3. Workflow

The DQEMG application does not currently propose any workflow or give any prompts or process information for the novice user. There is nothing, for instance, in the ordering of the screen buttons, to suggest that you want to evaluate the decomposition first, then edit particular sections of it, and then look at the results and summary screens and do your comparisons. This increases the amount of training and understanding necessary to use the software, but may make it more flexible for the researcher. However, clinicians involved in the Expert testing complained about it and requested more direction and transparent logic in the interface. Workflow and streamlining should be applied to the program in such a way that the experienced user does not lose functionality but the new and regular users are lead through the most typical steps of the process in a clear and efficient manner.

The first step in suggesting workflow and adding feedback to the user in DQEMG would be to have the screen following the MVC collection present a message to the user, such as "MVC successfully collected. Proceed to acquire contractions for the muscle study by hitting the **Acquire** button." The **Acquire** button could be highlighted with a light-colored outline at this point to make it easier to find. This would eliminate the situations observed in the expert user testing where the user completed the MVC but didn't know it and had no idea what to do next. Expert participants also requested feedback on this screen as to how high the baseline was measured to be, or how long in seconds the sample was, so the user has information with which to evaluate the quality of the MVC performed. The value of the baseline RMS is reported already, but the user has no way of evaluating that value. Some context or documentation could improve the situation.

The order of the buttons on the toolbar is a simple way to suggest workflow. In the right-hand top toolbar, **Decomp** should precede **Needle**, **Macro**, and **Ensemble**, with **Contraction Summary** following. To simplify the Decomposition screen, perhaps MUAPTs with fewer than 15 identified instances (which are automatically removed from the study by the decomposition application and are currently shown in gray) should be left off this screen altogether, as suggested by participant 4. A "Show excluded MUAPTs" option on the Options menu could be available for anyone who wants them to be shown again.

As part of making it clear what work has been completed and when it is time to move on, there should be some indication of whether or not a MUAP template or MUAPT has been assessed or edited by a human user. A checkbox for each MUAPT next to the train numbers on the Decomp screen, and check boxes in the upper left corners of the Micro template graphs on the Decomp and Needle screens would suffice. The Marker Editing screens could also provide an opportunity to check off the template graphs.

From the Decomp screen onward, every time a particular step is suggested by the common workflow, the button that would lead to that step could be highlighted, and the "Enter" key on the keyboard could be mapped to trigger the function of that highlighted button when it is hit by the user. When the user first arrives at the Decomp screen, the suggested function or button would be **Exclude**. When all MUAPTs have been reviewed and checked off, the default function then becomes **Markers**. On the Marker editing screen it may be useful for the default to be the **Offset** button until or unless the marker is checked off, in which case it becomes the **Next** button.

The Expert test participants demonstrated a preference for proceeding in an ascending direction through the numbered set of contractions, and wished to be able to edit all the marker landmarks in a study in an uninterrupted series. Therefore, the application should by default open

a new or archived study to the first contraction in the study, rather than the last. When the highest numbered MUAP template in a contraction is displayed in the marker editing screen, a **Next Contraction** button should appear in the lower right-hand corner of the screen, just above the toolbar. If this cannot be done then the **Close** button should be highlighted on that screen. There should also be a shortcut to move back to the first contraction, perhaps something like those discussed in Chapter 4.

It's possible that once the decompositions of all the contractions have been evaluated, it would make sense to show all of the MUAP templates on one needle screen and allow the user to edit them together seamlessly without having to distinguish between contractions. It should be possible to see them contraction by contraction, but participants also asked to be able to see an overview of all of them at the same time, and one participant commented that he didn't think he would even use the contraction summary, preferring to look at the results of the whole muscle study together.

The **Close** button should be renamed **Back** or **Return** and should always be in the same location on the screen. If this is the right-hand side of the screen, there should always be a right-justified toolbar to hold that button. Giving subscreens an outline or a pop-up window appearance could also make it more obvious to the user that they are in a subroutine and should hit a button to return to the parent area.

9.4.4. Modes; Visibility and Function

If the use of an "Edit" or "Exclude" mode and a "Markers" mode is retained, these modes should be made more visible. Visibility should especially be increased for the Exclude function. Also, the buttons for these modes should work like on/off switches. If the button is pushed once, the application will be in the mode and the button will be shown depressed. If the button is pushed again, it will release and the mode will be unselected. This will make it easier to browse through the information without being in a particular mode.

The Display mode buttons should be merged into a single button or toggle switch that switches from Display All mode to Valid Only mode. By default the Decomposition screen should have the Display All mode in effect until all the MUAPTs have been reviewed and checked off. Perhaps this button should be renamed Hide Excluded so that it either does or does not do the single function named, according to whether it is pushed in or not.

9.4.5. Muscle Study Results

9.4.5.1. More Information

Percent polyphasic MUAPs should be added to the results screen. Perhaps the range "min-max" of Micro MUAP values should also be given in the statistical table. An additional review screen that provides a look at all the MUAP waveforms that is accessible from the Results screen would also be useful.

9.4.5.2. Less Informational Clutter

Most macro information should be removed from the Results and Summary screens; the macro amplitude should be listed, followed by a button labeled "More Macro information" that opens a pop-up window with detailed information about the macro EMG. ID rate can be left off the Results screen. Area could also be removed. Furthermore the appropriate number of significant digits should be devised for each piece of data and that is all that should be displayed.

9.4.5.3. Characterization and History

Entering a characterization or comments on the study is a useful feature. The characterization input form should not be hidden within the function for adding the study to a database. If a diagnosis or study characterization has been entered, this should be accessible someplace. Not just what the characterization was, but also who made the characterization, why they characterized the study that way and the date the characterization was made. The date when the study was first collected and a history of changes made to it should also be accessible in the same location. The results screen may be an appropriate place for this.

9.4.6. Testing and Reliability

Proper testing requires a testing data set and a testing protocol that includes all the important tasks a user can logically be expected to attempt. Reliability issues for the DQEMG program are an identified barrier to the adoption of this program in clinical practice. Those issues include both the run-time stability of the application and the consistency of decomposition evaluation and DQEMG results characterization between different users.

Chapter 10 Conclusions

Quantitative EMG has been around for fifty years and could make a significant contribution to medicine. It is not commonly used by clinical physicians because the manual method takes far too much time and a computerized method has not yet been adopted in most hospitals. This study has identified a series of insights into the task of using the DQEMG application to collect and interpret quantitative EMG data for diagnostic purposes. Physician experts have been run through a user testing session and specific design suggestions have been made for improving the usability and efficiency of the DQEMG interface. To demonstrate how integrated information displays might improve the effectiveness of the DQEMG program, three modes of display were tested for performance and user preference. Principles from ecological interface design have informed our approach to this problem.

The system study, onsite observation and task analysis provided a detailed understanding of the transformation of information related to EMG, and how it leads to eliminating diagnostic possibilities, advising as to further testing, and monitoring trends in physiology in addition to direct diagnoses. A literature review identified a number of EMG characteristics that can be quantified and reported to the physician. Both the literature review and the entrance responses in the expert testing indicated that assessing MU recruitment can be crucial in distinguishing between disease processes. This study introduced the idea of reporting firing rate per MU in a QEMG application. Initial reaction in the Expert testing indicated FR/MU is at least a step in the right direction toward quantifying MU recruitment.

This study noted that since the physician is pressed for time, quantitative analysis needs to be reasonably quick to accomplish and interpret. More than one characteristic of the signal is necessary for clinical interpretation of QEMG data. Cognitive ergonomics suggests that information displays should take some of the mental workload off of the user by integrating multiple results into a display that can be interpreted quickly and easily.

The polar star display where all the axes are normalized so the standard deviation of normal is proportional between characteristics was demonstrated through the Nonexpert user testing to be an effective and popular solution to this problem. The polar star display produced the smallest number of errors in the Nonexpert testing and it was preferred by the majority of Nonexpert participants, though there were concerns about participant overconfidence on the polar star displays. The Nonexpert study indicated the histogram display that was used for the testing was not as usable. Participant comments indicate this may be partly due to the complexity of the

display – they described it as crowded with too much information they had to integrate. Any improved interpretability gained by adding normal reference lines to the histogram displays seemed to have been offset by the added complexity; results were inconclusive. A study using real data may be better able to assess the true contribution of a histogram display, with the presence of outliers that were missing in the simulated data. Amplitude, phases, and size index histograms are most likely to be useful according to this study and the analysis in chapter 3. There may also be a more effective range for the logarithmic scale for amplitude display. Designing the polar star plot specifically for the biceps brachii was possible in this study; more normative data would need to be collected to provide this type of graph for other muscles. As an expert test participant suggested, it would be best in the long run for each lab to be able to collect their own normative data and for the DQEMG application to provide a method of comparison to normative data for the same muscle.

The time required to perform and assess a quantitative EMG study was identified as another barrier to physician acceptance. From the Expert testing we conclude that the DQEMG program can be streamlined and made more usable for clinical purposes. Workflow hints in the interface and online user documentation will make it more clear to the user what procedure to follow and why. Clearer labeling and icon design and more effective color usage will make the interface easier to interpret and navigate. Providing a way to edit all of the micro templates in a study in one uninterrupted series, with aids to track what has and has not been reviewed, should reduce user frustration and increase acceptability. An improved offset identification algorithm combined with user training on how to place markers and how not to micro-edit would both be ways to dramatically reduce the amount of time needed for marker editing.

With these modifications as well as increased stability, the DQEMG application should be an acceptable and effective tool for EMG practitioners. Hopefully having a usable quantitative EMG program will further the application of combined indices such as size index and FR/MU. Research has demonstrated that these characteristics can improve discrimination between disease classes, but they are unavailable to the physician unless a computer-mediated quantitative analysis program like DQEMG is adopted. A trained physician and the background medical knowledge held by that physician will continue to be an integral part of any QEMG system. DQEMG may be able to increase the efficiency and the effectiveness of the physician's analysis of EMG data collected during a variety of muscle contraction levels. It would also be a valuable tool for research. Improved diagnosis of neuromuscular disease processes would provide a clear benefit to society. This application may also be used to help monitor disease processes and evaluate

treatments. Further research and development has been suggested that would make it even more effective for that task.

References

1. American Association of Electrodiagnostic Medicine: The Electrodiagnostic Evaluation of Patients with Suspected Cervical Radiculopathy: Literature Review on the Usefulness of Needle Electromyography. *Muscle & Nerve* 22: Supplement 8: S213-S221, 1999

2. Asano, Y., Yonemura, S., Hamada, H., & Ogawa, K: Method of ecological interface design applied to interactive diagnosis support system. In Y. Anzai, K. Ogawa, & H. Mori (Eds.), *Symbiosis of Human and Artifact,* 1995. (pp. 423-428). Amsterdam: Elsevier.

3. Basmajian JV, De Luca CJ: *Muscles Alive: Their Functions Revealed by Electromyography*, 5th Ed. Baltimore MD, William & Wilkins, 1985.

4. Bischoff C, Stålberg EV, Falck B, Edebol K, Olofson: Reference values of motor unit action potentials obtained with multi-MUAP analysis. *Muscle & Nerve* 1994;17:842-851.

5. Box G, Hunter W, Hunter JS: *Statistics for Experimenters*, John Wiley & Sons, 1978.

6. Buchthal F, Kamieniecka Z: The diagnostic yield of quantitative electromyography and quantified muscle biopsy in neuromuscular disorders. *Muscle & Nerve* 1982;5:265-280.

7. Croft, *Applied Statistics for Management Studies*. London, Macdonald and Evans LTD, 1979

8. DeSanctis G: Computer graphics as decision aids; directions for research. *Decision Sciences*, 1984,15:463-487.

9. Doherty T, Stashuk D: Decomposition-based quantitative electromyography (DQEMG): Methods and preliminary normative data. 2000.

10. Dumitru D, King JC: Nandedkar SD: Motor unit action potentials recorded with concentric electrodes: physiologic implications. *EEG Clin Neurophysiol* 1997a; 105: 333-339.

11. Ericsson, K.A., and Simon, H.: *Protocol Analysis: Verbal Reports as Data*. 1984 Cambridge, MA: MIT Press.

12. Gay A. EMG-based diagnosis of neuromuscular diseases: the clinical system, University of Waterloo, 2000 (unpublished).

13. Gibson J, *The Ecological Approach to Visual Perception,* © 1986 Lawrence Erlbaum Assoc. Inc.

14. Gruppen LD, Wolf FM, Billi JE: Information gathering and integration as sources of error in diagnostic decision making. *Med Decis Making* 1991;11:233-239

15. Hart S, Staveland L: Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research, *Human mental workload*. Hancock, Peter A. (Ed); Meshkati, Najmedin (Ed). 1988. Advances in psychology, 52 (pp. 139-183). Amsterdam, North-Holland.

16. Harvey, E: Correspondence by email, epharvey@uwaterloo.ca Statistical Consulting Office, University of Waterloo, Ontario 2001.

17. Herdmann J, Reiners K, Freund H-J: Motor unit recruitment order in neuropathic disease. *Electromyogr Clin Neurophysiol* 1988;28:53-60.

18. Howard JE, McGill KC, Dorfman LJ. Properties of motor unit action potentials recorded with concentric and monopolar needle electrodes; ADEMG analysis. *Muscle & Nerve* 1988b;11:1051-5.

19. Jaffe *Misused Statistics*. 1984

20. Jernigan, Fieguth *Introduction to Pattern Recognition*, 1999

21. Johnson EW, Fallon TJ, Wolfe C: Errors in EMG Reporting. *Arch Phys Med Rehabil* 1976;57:30-32.

22. Kimura J *Electrodiagnosis in Diseases* © 1989.

23. Lambert E. Electrodiagnosis of neuromuscular disease, *Pediatrics* 1964; 34:599-600.

24. Matthews DE, Farewell VT. *Using and Understanding Medical Statistics*, © 1996 S. Kargar AG.

25. Mendenhall W, Reinmuth J: *Statistics for Management and Economics*, 3rd Edition, Duxbury Press 1978.

26. Neilsen, J: *Usability Engineering,* ©1993 Academic Press.

27. Nirkko A, Rosler K, Hess C: Sensitivity and specificity of needle electromyography: a prospective study comparing automated interference pattern analysis with single motor unit potential analysis. *EEG Clin Neurophysiol* 1995;97:1-10.

28. Okajima Y, Tomita Y, Sasa H, Tanaka N, Kimura A, Chino N: The size index as a motor unit identifier in electromyography examined by numerical calculation. *J Electromyography and Kinesiology* 1999;9:201-208

29. Pfieffer G, Kunze K: Discriminant classification of motor unit potentials (MUPs) successfully separates neurogenic and myopathic conditions. *EEG Clin Neurophysiol* 1995;97:191-207.

30. Preston DC, Shapiro BE. *Electromyography and Neuromuscular Disorders; Clinical-Electrophysiologic Correlations*, Butterworth-Heinemann, 1998

31. Smith P. Debunking the myths of UI design IBM, Toronto Software Lab, March 2001.

http://www-106.ibm.com/developerworks/library/us-myth.html?dwzone=usability

32. Sonoo M, Stålberg E: The ability of MUP parameters to discriminate between normal and neurogenic MUPs in concentric needle EMG: analysis of the MUP "thickness" and the proposal of "size index." *EEG Clin Neurophysiol* 1993; 89:291-303.

33. Stålberg E, Andreassen S, Falck B, Lang H, Rosenfalck A, Trojaborg: Quantitative analysis of individual motor unit potentials: a proposition for standardized terminology and criteria for measurement. *J. Clin Neurophysiol* 1986;3:313-348.

34. Stålberg E, Bischoff C, Falck B. Outliers, a way to detect abnormality in quantitative EMG. *Muscle & Nerve* 1994;17:392-399.

35. Stålberg E, Falck B: The role of electromyography in neurology. *EEG Clin Neurophysiol* 1997; 103: 579-598.

36. Stålberg E, Falck B, Sonoo M, Stalberg S, Astrom M: Multi-MUP EMG analysis - a two year experience in daily clinical work. *EEG Clin Neurophysiol* 1995;97:145-154.

37. Stålberg E, Nandedkar SD, Sanders DB, Falk B. Quantitative Motor Unit Potential Analysis. *J. Clin Neurophysiol* 1996; 13(5): 401-422

38. Stashuk D: Decomposition and quantitative analysis of clinical electromyographic signals. *Med Eng & Physics* 1999a; 21: 389-404.

39. Trujillo A, Schutte P: Non-Traditional Displays for Mission Monitoring. ANS Conference Abstract LOG NO. 056 12/01/99. http://techreports.larc.nasa.gov/ltrs/PDF/1999/mtg/NASA-99-answm-act.pdf

40. Vicente K: Memory recall in a process control system: A measure of expertise and display effectiveness. *Memory & Cognition* 1992; 20 (4), 356-373.

41. Wainer H: How to Display Data Badly, *The American Statistician*, May 1984, 38 (2): 137-147.

42. Wickens CD: *Engineering Psychology and Human Performance, second edition*. 1992 New York: HarperCollins Publishers.

43. Winer B. J. *Statistical Principles in Experimental Design,* McGraw-Hill, 1971.

44. Zalewska E, Hausmanowa-Petrusewicz I: Global and detailed features of motor unit potentials in myogenic and neurogenic disorders. *Med Eng & Physics* 1999;21:421-429.

45. Zalewska E, Hausmanowa-Petrusewicz I: Effectiveness of motor unit potentials classification using various parameters and indexes. *Clin Neurophysiol* 2000;111:1380-1387