ABSTRACT

WAVELET TRANSFORM METHODS FOR IDENTIFYING ONSET OF SEMG ACTIVITY

by
Janina Wilen

Quantifying improvements in motor control is predicated on the accurate identification of the onset of surface electromyograpic (sEMG) activity. Applying methods from wavelet theory developed in the past decade to digitized signals, a robust algorithm has been designed for use with sEMG collected during reaching tasks executed with the less-affected arm of stroke patients. The method applied both Discretized Continuous Wavelet Transforms (CWT) and Discrete Wavelet Transforms (DWT) for event detection and no-lag filtering, respectively. Input parameters were extracted from the assessed signals.

The onset times found in the sEMG signals using the wavelet method were compared with physiological instants of motion onset, determined from video data. Robustness was evaluated by considering the response in onset time with variations of input parameter values.

The wavelet method found physiologically relevant onset times in all signals, averaging 147 ms prior to motion onset, compared to predicted onset latencies of 90-110 ms. Latency exhibited slight dependence on subject, but no other variables.

WAVELET TRANSFORM METHODS FOR IDENTIFYING ONSET OF SEMG ACTIVITY

by
Janina Wilen

A Master's Thesis
Submitted to The Faculty of
New Jersey Institute of Technology
In Partial Fulfillment of the Requirements for the Degree of
Master of Science in Biomedical Engineering

Department of Biomedical Engineering

January 2004

APPROVAL PAGE

WAVELET TRANSFORM METHODS FOR IDENTIFYING ONSET OF SEMG ACTIVITY

Janina Wilen

Dr. Stanley Reisman **Date**
Professor and Graduate Advisor of Biomedical Engineering, NJIT

Dr. Richard Foulds, Committee Member **Date**
Associate Professor of Biomedical Engineering, NJIT

Dr. Gail Forrest, Committee Member **Date**
Clinical Research Scientist, Kessler Medical Rehabilitation Research and Education Corporation, Human Performance and Movement Analysis Laboratory

---

Dr. Xiaorui Tang, Committee Member **Date**
Research Scientist, Kessler Medical Rehabilitation Research and Education Corporation, Rehabilitation Engineering and Analysis Laboratory

BIOGRAPHICAL SKETCH

Author: Janina Wilen

Degree: Master of Science

Date: January 2004

Undergraduate and Graduate Education:

- Master of Science in Biomedical Engineering
  New Jersey Institute of Technology, Newark, NJ, 2004

- Bachelor of Science in Biomedical Engineering
  Cornell University, Ithaca, NY, 1999

Major: Biomedical Engineering

Presentations and Publications:

Wilen, J., Sisto, S. A., & Kirshblum, S. (2002). Algorithm for the detection of muscle activation in surface electromyograms during periodic motion. Annals of Biomedical Engineering, 30(1), 97-106.

Wilen, J., Sisto, SA., & Kirshblum, S. (2001). Algorithm for the Detection of Activation/Deactivation Instances in Continous EMG Signals. 2001 Annual Fall Meeting of the Biomedical Engineering Society (S-16).

Wilen, J, Sisto, SA., Kirshblum, SC. (2001). Reliability of an Algorithm for the Detection of Muscle Activation in Surface Electromyograms During Periodic Wheelchair Propulsion. (Abstract) Gait and Posture, 13(3):283. Presented at Gait and Clinical Movement Analysis Society Meeting, Sacramento, CA, April 2001.

"On considering the Divine acts, or the processes of Nature, we get an insight into the prudence and wisdom of God as displayed in the creation of beings, with the gradual development of the movements of their limbs and the relative positions of the latter...The nerves are enabled to set the limbs in motion... I quote this one instance because it is the most evident of...wonders."

Rabbi Moshe Ben Maimon (Maimonidies), Guide to the Perplexed, Part III, Chapter XXXII, Jewish Scholar and Physician to the Caliph of Egypt, 1135-1204

To my father who taught me to love learning,
my mother who trained me to be an unintentional feminist,
and to both, who are kvelling.

To my sister, Debbie who I look up to always
and to my twin sister, Arielle who went down and up with me

To Elyana who told me I do too much homework.

And to all of us who draw water in the desert

ACKNOWLEDGMENTS

Foremost, gratitude goes to Dr. Stanley Reisman who served as my thesis advisor, and also as teacher throughout all my years at NJIT as a biomedical engineering student. His combined skepticism and faith in my abilities pushed me farther than I was able to recognize.

Thanks are given to Dr. Gail Forrest for lending me her knowledge, technical advice as well as feedback in the middle of a stressful workload. Her insights were not only contextual, but practical as well.

Dr. Xiaouri Tang was instrumental in providing guidance, by asking all the right questions, and keeping me on track.

Appreciation goes to Dr. Richard Foulds whose exuberant support of my love of sharing knowledge gave me great hope.

The research was done under the umbrella of Kessler Medical Research Rehabilitation and Education Corporation, in West Orange New Jersey. The data was made available to me by a grant from the Retirement Research Foundation, under a study headed by Dr. Stephen J. Page.

I would like to extend a tremendous thank you to Dr. Sue Ann Sisto, my supervisor at KMRREC who asked me to do things I thought far over my head since the day I began, without whom I would not have thought to try, and who laughed when I said "I don't know how".

A special note of appreciation must be handed to Anne Marie Petrock, a giver and fellow student, who showed me how to do this right.

TABLE OF CONTENTS

TABLE OF CONTENTS
(Continued)

TABLE OF CONTENTS
(Continued)

Chapter **Page**

LIST OF TABLES

LIST OF FIGURES

LIST OF FIGURES
(Continued)

LIST OF FIGURES
(Continued)

Figure **Page**

LIST OF FIGURES
(Continued)

Figure **Page**

CHAPTER 1

INTRODUCTION

1.1 Objective

The objective of this paper is to demonstrate the design of a wavelet transform method for the accurate discrimination of muscle activation in surface electromyography (sEMG) signals. For the development of the method, sEMG signals from the less-affected biceps brachii of stroke patients have been used.

"The myoelectric signal is the electrical manifestation of the neuromuscular activation associated with a contracting muscle" [31]. Recorded since the 1920's [42], electromyography (EMG) signals have been found to be very useful in understanding the relationship between a motion and the muscles involved in generating the motion. EMG displays patterns of motor unit activity, and can be used to quantify the effect of neurological disorders on movement, as well as to document motor behavior in conjunction with kinematic data. Surface EMG (sEMG), the type collected in this study, involves recording the signal from simple electrodes during muscle contraction, that are strategically placed on the skin surface near the main muscle mass. The recording of that signal is strongly influenced by the anatomical and physiological properties of that muscle, what motion is being recorded, and the properties of the recording system. The result is a very complex signal.

From a medical perspective, understanding these periods of activity can help define when muscles are expending unnecessary energy, possibly countering the effect of agonist muscles [102]. In kinesiology and kinematic studies, determining the precise timing of these EMG and sEMG events is primary, particularly when activity patterns of

antagonist muscle pairs are being considered [99]. Detection of activation in a signal is predicated on the fact that there is a linear, or near-linear, association between the amplitude of an EMG signal during isometric contractions and the tension produced by that muscle [18]. Tension refers to that which is generated along the line of the muscle fiber as a result of contracting fibers, regardless of the effect on limb position, and therefore, may be isometric. In contrast, force implies a directional vector applied to an object against inertial properties of the object, such as that required to lift a limb. Any force generation of the muscle is preceded by a buildup of tension in the muscle, evidenced in EMG. It is the beginning of tension for the purpose of force generation that is being studied here. Determining temporal patterns in EMG requires a tool for finding the start and end of muscle activation [102].

Methods for determining point of onset of activity have long been under development. Thresholding, statistical counting and predictions, pattern matching, and even neural networking and fuzzy logic methods have been applied. Still, these are no more accurate than visual detection [21; 34]. These methods are also often dependent on external or secondary sets of information for their application, such as maximum voluntary contractions, noise or resting signals. Furthermore, significant alteration of the original signal is needed to allow those methods to function.

Wavelet transforms (WT) are a relatively new signal processing technique different from other methods, in that it considers the content of the complex signal at each instant, independent of topological patterns. In contrast to other methods, a WT can use a signal's own characteristics of amplitude and instantaneous frequency for defining appropriate analysis, instead of using a second signal to represent probable noise or

maximum amplitude of the signal of interest. The non-periodic EMG signal is an obvious application for using this new analysis technique of wavelet analysis for accurate feature extraction [96]. A quick and robust method for using WTs for onset detection has been developed here.

To demonstrate the viability of the detected onsets, latencies between the identified onset times in sEMG and initiation of motion in the less-affected arm of stroke patients was compared to latencies reported in the literature. Robustness of the method was evidenced by the non-critical response of measured latencies from highly variable inputs.

1.2 Background

1.2.1 Physiology of Electromyography

Electromyography is the recording of the wave of the electrochemical stimulus required to contract a muscle. The motor unit (MU) is the smallest controllable unit of the muscle, each separately controlled by a motor neuron axon. A MU consists of a synaptic junction in the ventral root of the spinal cord, a motor axon, and a motor end plate or junction in several muscle fibers. One motor unit (MU) controls between 3 and 2000 muscle fibers, depending on the fineness of required control [97].

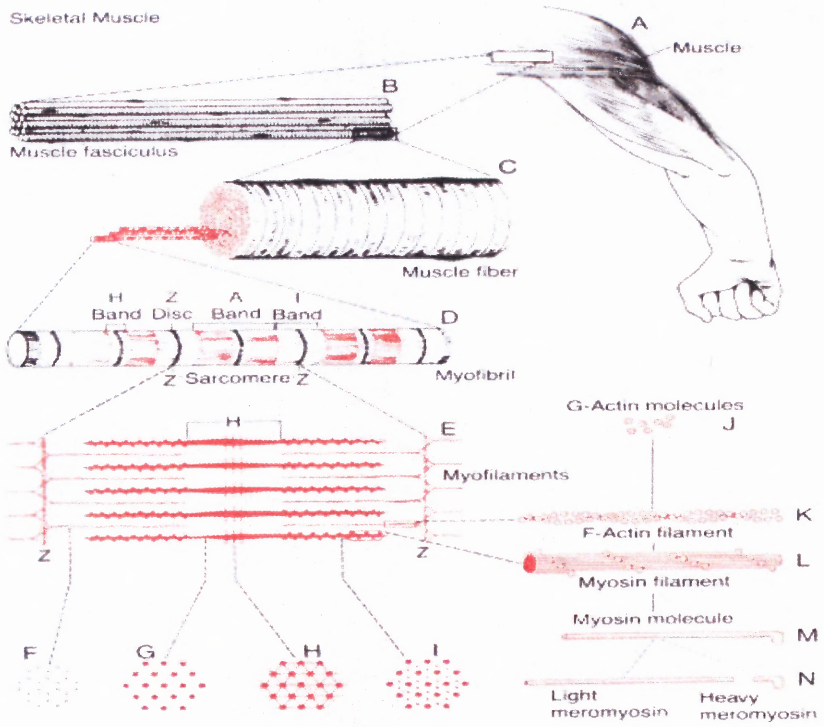

Figure 1.1 Organization of skeletal muscle.
Source: Guyton 1996

A muscle fiber, C in Figure 1.1, is about 100 µm in diameter, and is composed of several thousand chains of contractile elements, each element called a sarcomere. The fiber is the unit directly stimulated by a motor neuron. Within a fiber are 1 µm myofibrils, in which there are filaments of about 100 Ä [97].

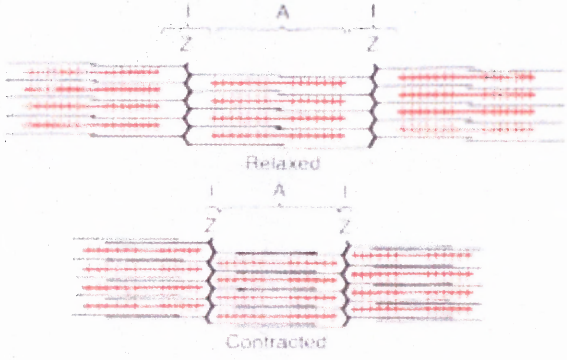

Figure 1.2 The dense areas in myofibrils that mark the beginning of sarcomeres are z-lines. The actin filaments of the sarcomeres are anchored in the Z lines.
Source: Guyton 1996

In the fibrils (D in Figure 1.1), two types of protein bands, actin and myosin, are interlaced in units called sarcomeres (Figure 1.2). The interaction between the two proteins causes contraction. The space between them has crossbridges that flex to create tension, and release to lengthen.

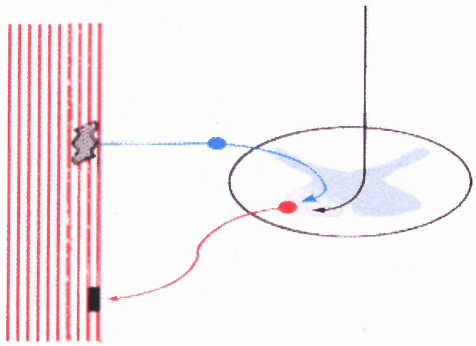

Figure 1.3 The reflex arc. A signal comes out the spinal cord, and then is moderated by sensors at the muscle fiber.

Muscle fibers are activated after the control signal is passed through and moderated by the reflex arc (Figure 1.3). Out the anterior root of the cord, the signal passes into the anterior motor neurons, that become both alpha and gamma motor neurons. Two sensory units, the muscle spindle fibers within the muscle, and golgi tendon organs at the junction of muscle and bone, provide feedback for muscle control through interneurons, and to the brain through the spinocerebellar tracts. Stimulation of a single alpha neuron fiber excites a motor unit, but the excitation is moderated by the gamma neuron. The gamma neuron transmits the same signal to the intrafusal fibers in the muscle spindle, responsible for detecting rate of stretch of the muscle fiber. In response to stretch, the spindle increases firing rate to the interneurons, and decreases the rate in response to muscle shortening. The return loop to the alpha and gamma neurons

occurs through the interneurons, thus moderating the actual input to the fiber from the alpha neurons [50].

A Motor Unit Action Potential (MUAP) is the conduction of an electrical potential in a muscle fiber as a result of the recruitment of the MU. A MUAP is approximately 100 μV in amplitude and 4 m/s in velocity [123].

It is in the sarcomeres that the following reaction occurs with the onset of a stimulus. The innervation of the motor unit by the motor neuron is transmitted across a synaptic junction called a motor end plate. An action potential (AP) travels down the motor neuron and triggers electrochemical reactions at the junction. ACh is released, crossing the gap (about 200-500 Angstrom units across) and, if the voltage potential crosses a threshold of about 70mV, causes a depolarization in the form of another AP wave in the post-synaptic membrane of the muscle fiber (Figure 1.4). It is this AP stimulus that reaches the Z line of the sarcomere, creating an AP in the muscle fiber membrane. A single wave through a muscle fiber is a MUAP [123]. When the AP spreads through fibers fast enough (2-5 m/s) to contract all the sarcomeres at once, there is a tension to generate or inhibit motion [92].

A single MUAP, starting at the z-line, spreads a stimulus along the transverse tubular system of a motor nerve of the contractile element in the neuron, releasing Ca^{++} into the sarcoplasmic reticulum. The Ca^{++} diffuses into the actin and myosin filaments in the muscle fibers, thus hydrolyzing ATP and generating mechanical energy (along with ADP), in the form of contractile forces at the crossbridges of the actin and myosin filaments.

The flood of ionic charge carried by these ions into and back out of the neuron is modeled as a depolarization wave in time, along the direction of the muscle fibers, as in Figure 1.4.

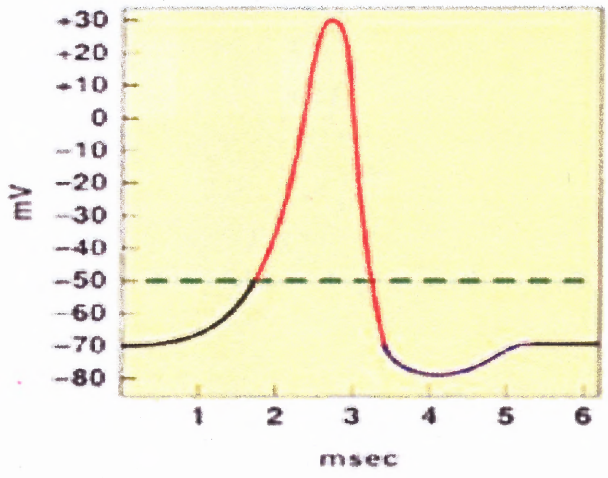

Figure 1.4 A MUAP depolarization wave. It is this voltage change that is measured by electrodes.

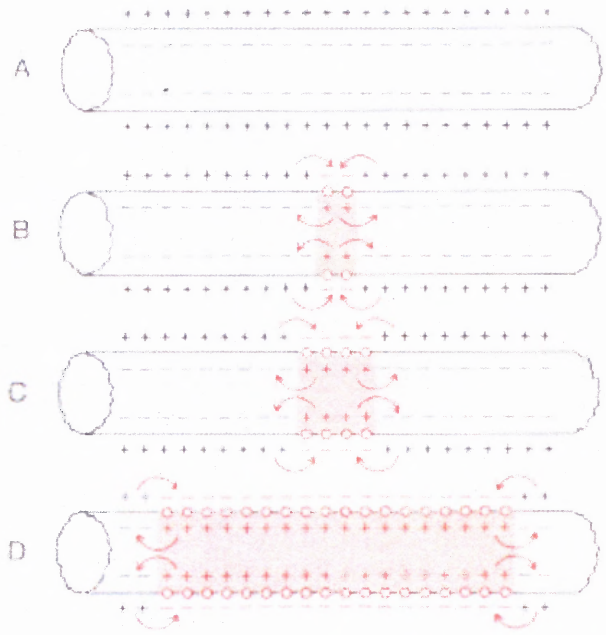

Figure 1.5 Propagation of action potentials in both directions along a muscle fiber. The difference in charge (between + and -) constitutes a voltage difference, and that is the front of the propagation wave.
(Source: Guyton) [50]

The time excursion of the voltage potential with respect to ground caused by the flow of charged ions is what is termed an action potential (AP) [31]. The depolarization wave front (Figure 1.5) and the subsequent repolarization of the fiber is detected by electrodes, either at the skin surface or subcutaneously near the motor unit.

Berardelli [8] stated that all input to the MU afferent of the spinal cord will not directly induce instantaneous rises and falls in the motor unit firing and the EMG recording. The height of the pulse into the motor neuron pool controls the rate of EMG rise, but not how soon a MU will respond.

Excitatory input to motor neurons have finite maximums of about 70 ms. Thus, at a neuromuscular level, usually recorded with fine wire, EMG bursts of MU activity also have durations averaging 70-80 ms, within a range of 16-85 ms [46]. A MU twitch in the biceps bracchii lasts about 52 ms. In a standard 70 ms burst, maximum tension is only produced with a train of at least 5 – 10 impulses per motor unit. Lengthening input pulses longer than 70 ms allows the summation of input pulses for the production of tension sufficient to generate motion, or force generation. In this way, the peak tension in an isometric motion occurs after 100-150 ms after onset of EMG excitation. Recruitment of input pulses stops occurring in a MU at approximately 75% of the maximum voluntary isometric tension of most muscles. The biceps brachii, however, stops at 60% [31].

Each muscle has a finite number of motor units, each of which is controlled by a separate nerve ending. If enough MUAPs come along and last long enough, the MU is activated and responds with a mechanical response called a twitch. Provided the frequency of incoming MUAPs is sufficient, with a delay of a few ms, a MUAP would be accompanied by a muscle fiber twitch. An increase in tension in a muscle can be

accomplished either by increasing the stimulation rate into one MU, or by recruiting other MUs in the muscle [123]. A twitch becomes a tension generating muscle contraction when the number of MUs recruited at once reaches 10% of that muscle's recruitment at maximum voluntary contraction [42]. Initially, the whole muscle increases tension with increased MUAP firing rates, supplemented by adding newly activated MUs [31; 97]. Each MU has its own time course of the twitch, of the maintenance of tension.

The firing rate, or the frequency of firing, contributes to tension in one MU. Therefore, both in its amplitude and frequency content, recorded EMG reflects the number of motor units recruited for a contraction [92].

Smaller motor neurons are activated first during voluntary contraction, because the number of muscle fibers innervated by one motor neuron is directly related to the size of the innervating neuron (small neuron, small area of contact) [31; 123]. Smaller motor units require less stimulation to activate, because they have slower conduction velocities, and lower excitation thresholds. They also remain activated longer – they are slow twitch. Thus, MUAPs recorded at lower tension levels are of the smaller variety. They have longer contraction times, but produce lower twitch tension. The reverse is true for larger motor units – they are fast twitch fibers [31], which produce greater isometric tension. The neuromuscular delay between initial MUAP recording and the initiation of a motion produced by the biceps brachii, would be larger than one recorded in the biceps femoris – a large thigh muscle. Thus, onset of activity in gait would be faster, with quicker rises to peak, than the data tested here, which were signals recorded from the biceps brachii.

Hudgins [38] demonstrated that a significant amount of information is present at least 100 ms prior to a motion in normals, with widely spaced electrodes. It is in those

100 ms after the initial MUAPs are received that recordable muscle recruitment occurs, but there is not yet enough recruitment for the generated tension to initiate motion.

1.2.2 Surface Electromyography

Electromyography (EMG) is the recorded electronic signals of multiple MUAP trains from active motor unit fibers occurring at the same time, underneath the electrode, that instruct a muscle to contract [127]. The electrodes pick up the electromagnetic field accompanying the movement of charged ions that is recorded as a voltage change [108]. Since muscle fibers of a motor unit are randomly distributed through a subsection of muscle, a muscle recording is intermingled with fibers of different sizes belonging to different motor units [32]. Thus, even embedded fine-wire electrodes may read across 20-50 motor units, never single MUAPs.

Figure 1.6 maps many of the influences to the sEMG pattern at onset of tension. Some of these factors are fiber size type and health, firing rate, duration of firing, degree of overlap, as well as the mechanics of the motion being committed, the amount of hydration of the skin, the amount of fat tissue between the skin and the muscle being recorded, and various electrical properties of the electrode and recording instruments [123]. It is important to note that the complex signal is influenced by frequencies, firing rates, and speeds of incoming signals, as well as their magnitudes.

Furthermore, the low pass filtering properties of the skin/electrode combination may introduce spectral shifts in recorded MUAPs, in the form of lengthening durations and phases [32]. Larger MUs produce more isometric tension. MUs with increased recruitment thresholds and twitch tension have increased MUAP amplitudes. Thus, the

amplitude of the pattern seen in the EMG is as much connected to the size of the muscle fiber as to the number recruited MUs.

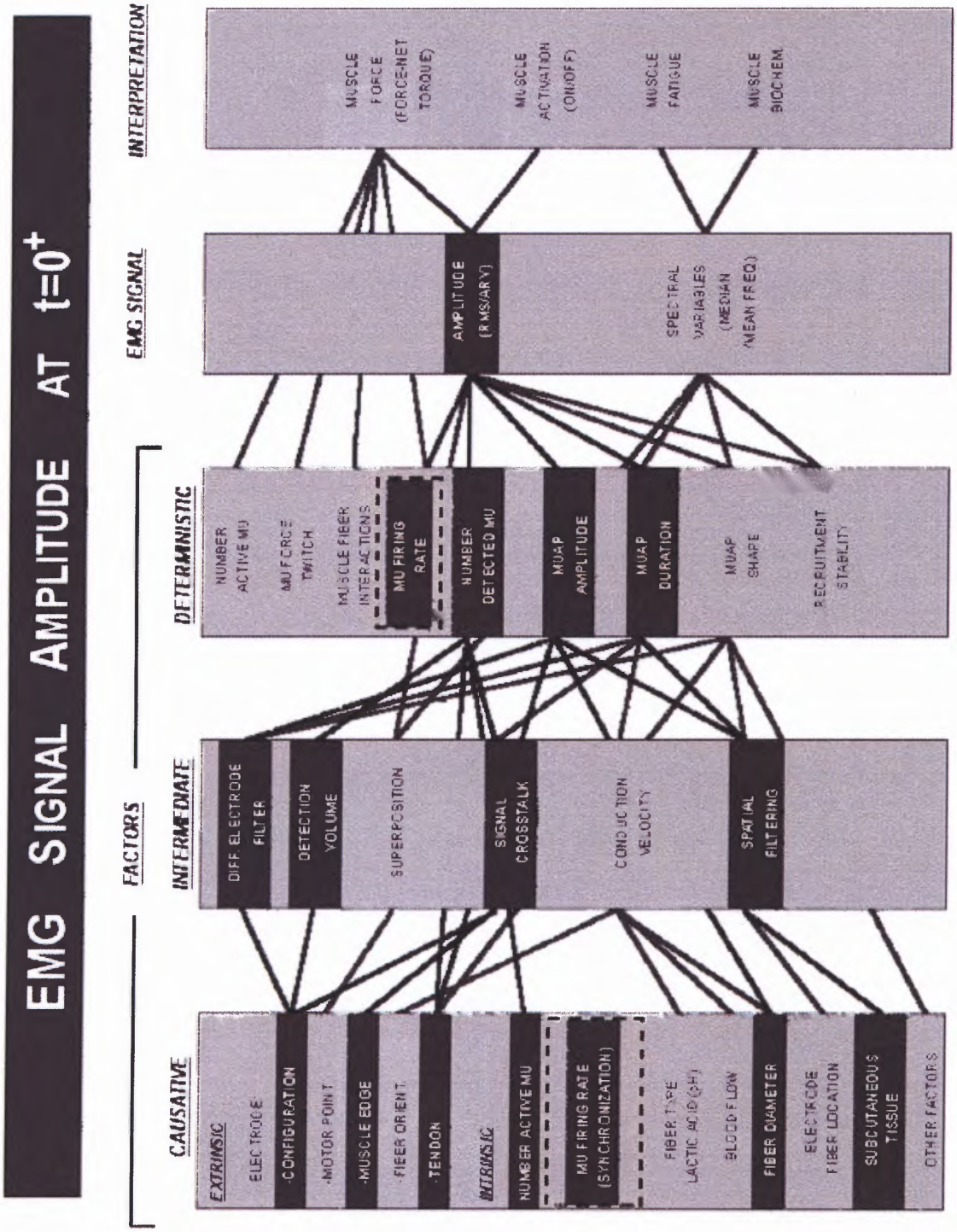

Figure 1.6 There are many factors influencing EMG during activation. They are both amplitude and frequency related.
(Source: Deluca 1997) [32]

An interference pattern (IP) is a waveform of the summed MUAPs from a number of muscle fibers, and is the form MUAP recordings take when surface electrodes are used [4]. Signals recorded with indwelling fine wires are different from the IP recorded with surface electrodes for sEMG. The actual MUAP is only 1-9% of the amplitude of the MVC, but the IP is 10-100% of MVC amplitude [42]. Because of the large surface area covered by a skin surface electrode, the shape of the recorded MUAP IP in sEMG no longer accurately reflects the actual MUAP innervation going on under the electrode [127].

The relationship between sEMG and tension generation is still under discussion [99]. Both amplitude and firing rate should be taken into account. A standard method for assessing muscle fatigue is to measure the change in mean frequency. Muscle fatigue is the reduction in the maximum tension the patient is able to generate, and may show up as a reduction of high frequency content in the signal, but not as amplitude changes. Despite both amplitude and frequency being characterizing changes, because firing rate and actual number of firing motor units is not easily extracted information from sEMG, amplitude and topological patterns are often used as the sole indicator of activity. Furthermore, the magnitude of sEMG becomes an estimated parameter after smoothing with traditional methods including rectification, linear enveloping, integration, and RMS calculation. Wavelets consider changes in the time domain alongside changes in the frequency domain. Wavelets can be used for identifying the point in time of the shift of a muscle into a fatigued state, or the beginning of a slow rate of increase of EMG activity related to force generation [96].

For the tasks studied in this paper, two disposable electrodes were placed on the upper arm: one on the main belly of the biceps brachii, the other on a nearby bony landmark. A third body reference electrode was pasted to the back of the subject. Ideally, the first two electrodes should be set up for differential amplification, with both electrodes just over the belly of the muscle, about 2 cm apart. The differential setup subtracts the noise from both electrode inputs, thus:

$$(S1 + \text{noise}) - (S2 + \text{noise}) = S1 - S2. \tag{1.1}$$

Due to the distance between the electrodes, the signal input to both is at a slightly different point in the time path of the MUAP. The time it takes for an AP wavefront to cross that distance is significant on a neurological level. The time lag results in a recordable voltage difference, which is the MUAP wavefront seen in Figures 1.4 and 1.5. Thus, whatever part of both captured signals is the same, is global noise, and is removed when one signal is subtracted from the other. Ideally, the difference relating to the movement of charge along the muscle fiber is the only displayed content. This arrangement serves as a differential amplifier, because it is the change along the spatial line of the muscle fiber that is a MUAP. Thus, if an electrode is over a region with negative charge, and the other electrode is over a positively charged region, the difference in voltage between the two electrodes is captured [32].

Power spectrum analysis is a standard tool of decomposing sEMG, and its mathematics will be addressed further (see Subsection 2.4.1). It is useful when studying EMG signals on a gross scale, and when determining frequency bands of interest. The major frequency band of the power spectrum is due to the MUAP [42]. Basmajian, in 1970, established normal durations recordable by surface electrodes, of about 6-60 ms, or

50-150 Hz [58]. Since then, the frequency band of the EMG signal has been found to be concentrated between 15.6-250 Hz [27; 97], but has content up to 1000 Hz. The bandlimit of the captured sEMG signal was from 10-1200 Hz, and digitally sampled at twice that, at 2400 Hz, in consideration of the Nyquist theorem. The theorem states that, to avoid frequency aliasing, a signal must be captured at a sampling rate greater than twice the maximum frequency present in the signal.

Firing frequencies, when MUAPs themselves are recordable, can affect the power spectrum when they are between 10-30 Hz. Higher frequencies are short duration and rapid rise polyphasic waveforms. Low frequencies are similarly long duration MUAPS with slow-rise edges. The IP in sEMG is reflective of MUAP characteristics.

The spacing of the electrodes also affects the way sEMG records MUAP IP patterns. Because the propagation wavefront of a MUAP must cross the distance between two electrodes, the farther apart they are, the slower traveling, thus longer, the MUAP will seem to be. Signals recorded from more closely spaced electrodes will have higher frequency components. Similarly, surface electrodes record longer MUAPs than indwelling wire electrodes, lasting 6 – 40 ms compared to 3 – 20 ms [123], because they are farther apart, in addition to the fact that they receive signals filtered by the skin, and read IPs of multiple MUAPs, instead of unitary MUAPs. If one surface electrode is placed too close to an innervation zone, as in Figure 1.7, the electrode may record MUAP signals traveling in both directions along the muscle (which occurs at the start point of the MUAP, seen in Figure 1.5), which doubles the maximum frequency found in the signal as in Figure 1.8. However, if they are placed too closely, they may filter out high frequency information. If they are too far, aliasing may occur [32].

It is relevant to note that the electrode placement for the reach data was not the best, as apparent in Deluca 1997's recommendation for electrode placement. Indeed, the power spectrums for the sEMG of these data appeared similar to those demonstrated in Figure 1.7. Figure 1.8 demonstrates frequencies up to 400 Hz, reflective of poor electrode placement. Due to the expiration of the study grant and time limits, the placement error was not corrected, and a signal recorded from properly placed electrodes were unavailable.

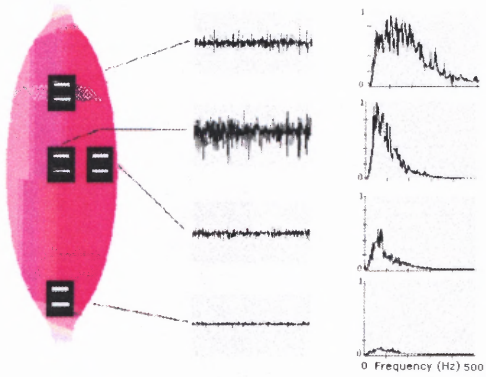

Figure 1.7 Optimal electrode placement is achieved with the center electrode pair. The top right plot resembles the power spectrums calculated with this data, and represents electrodes placed too close to the innervation zone.
(Source: Deluca 1997) [32]

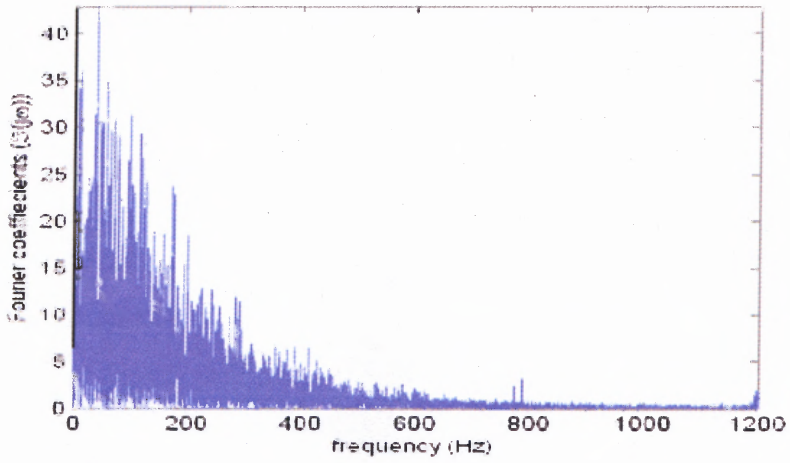

Figure 1.8 Sample power spectrum calculated with subject two, trial three, pre training. Content is visible into 400 Hz, similar to the example in the upper right plot of Figure 1.7.

In a clinical scenario, the deciphering of the activation and deactivation events is commonly done on EMG collected with surface electrodes, due to their noninvasive nature, even though indwelling wire electrodes can better target MUs for individual MUAP recording.

In sEMG signals, noise other than physiologically relevant content can appear. Low frequency noise (5 –15 Hz) can be from movement (10 Hz according to [123]), called movement artifact. Movement artifact looks like a baseline shift of the entire signal. It is actually a result of two electrodes moving closer and farther with motion, causing a change in distance and impedance between the two, which affects the delay between each electrode's recording of the same information. A second major source of error is a 60 Hz sine wave (in the USA; 50 Hz in Europe) that can show up from surrounding electronic units [123].

1.2.3 Background of Stroke Rehabilitation

Gillen [47] provides a background into the cause of a cerebrovascular accident (CVA), or stroke. A stroke occurs when the blood flow carrying oxygen to the brain is blocked in the vessel, either by an obstruction or by a breakage of the blood vessels themselves. These two pathologies are ischemic (the more common type) and hemorrhagic, respectively. Ischemic stroke occurs either when the cardiac output falters, such as with cardiac arrhythmias or acute infarctions in the heart, consequently stopping transport of oxygen to the brain. A cardiac embolism can also occlude the arteries to the brain.

Vascular causes of hemorrhagic strokes are less common, and are usually due to plaque in the walls of the aorta, carotid arteries, or smaller vessels in the cerebral circulation. Plaque formation usually occurs at the branching points of major vessels, and

may be followed by damage to the inner wall. In response to the damage, swelling and arteriosclerosis occurs, and then the vessel calcifies, narrowing the passage and creating turbulent flow. It is here that the plaque can break off to a thrombosis. A thrombotic stroke is usually related to abnormalities in the arterial wall.

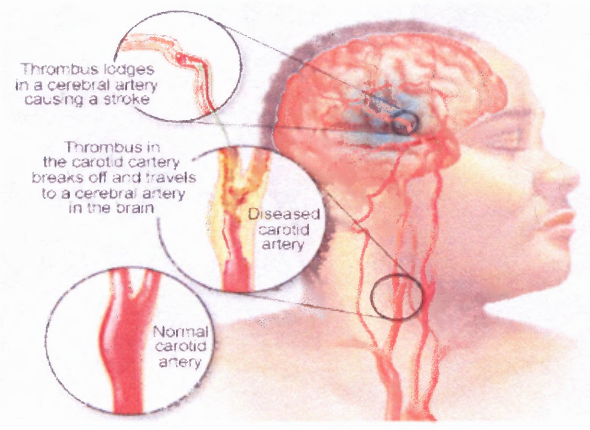

Figure 1.9 A thrombotic hemorrhagic stroke.
(Source: Guide to ischemic stroke; provided by Mayo Foundation for Medicine 1998-2003)

Abnormalities include dissection of the vessel, external compression on the vessel, or hematological disorders. Both thrombosis and embolism are usually present.

Stroke is the third leading cause of mortality in the US, with three million victims who survive yearly. However, those survivors are often afflicted with syndromes including aphasia, pain, sensory loss, memory deficits, and loss of motor skills. Any CVA occurring in the main cerebral arteries results in contralateral hemiplegia. A Copenhagen study in 1994 found that 69% of stroke patients had mild to severe upper extremity dysfunction [47]. After the acute stage, the rehabilitative task is to retrain the patient to function with the good arm as compensatory, as well as to regain as many motor skills in the paretic arm as possible. Still, Wade 1983 [20] reported that only 20% of patients with flaccid upper limbs two weeks after stroke ever regained functional use

of their hand. More recent clinical studies have demonstrated improvements even one year later [20; 119].

The basis of retraining requires applying functional tasks – not just arbitrary motions. The fact that the patient may be unable to perform functional tasks such as grasping, self-motorization or directional control, is a common confound. The reaching task is a necessary training stage before grasping or lifting can be achieved [20].

A concrete goal on the part of the patient is to be able to really use their arm, not just move it better. Rehabilitation should be designed to help a patient accomplish their goals [47]. The reaching task as a retraining task is valuable because it addresses multiple levels of control, and because it is a functional ability.

Carr [20] one possible delineation of the stages involved in a reaching motor task. First, 1) the basal ganglia scales the size of initiating agonist bursts by way of voluntary command and inhibition of inappropriate EMG. 2) The primary cortex and premotor areas of the cerebellum moderate selection of motor strategy by controlling timing in those voluntary bursts, increasing muscle tension phasically. 3) Next, the corticospinal tract becomes involved on a neuromuscular level, and determines spatial and temporal recruitment of motor units. 4) Proprioceptive feedback into the reflex arc (Figure 1.3) contributes to the accuracy of the trajectory and the end point [8]. The final stage of a force generating motor task is MU activation.

The cortical level, stages one and two, include conceptualizing the activity as well as afferent control. A stroke, or cerebrovascular accident (CVA) will not affect all four stages. Only the cerebellar portion of voluntary control moderation in stages one and two is affected, leaving actual innervation of the muscle unaffected. Paresis is due to

impairments in first two stages of a voluntary motor task. Final motor output in stroke patients can be additionally moderated by changes in stage three, by moderation of descending and propriospinal excitatory and inhibitory inputs into the spinal interneuron and alpha motor neurons in the reflex arc [21]. Stage three is all extra-cortical, is not directly affected by the CVA, and is the activation that is detected with sEMG.

The physiognomy of a reaching task is as follows: In a seated position, postural muscles stabilize the trunk. For an optimal reaching task, the shoulder is horizontally flexed, and the elbow extends to 180 degrees. The biceps are involved in shoulder flexion and act as antagonists to the triceps during elbow flexion. The scapula protracts and elevates, and, if functional in a CVA patient, the rotator cuff keeps the shoulder in the socket. The elbow extends with extensor activation including the triceps, and the biceps brachii help stabilize antagonistically at the elbow.

Confounds to a CVA victim's ability to perform a reaching task are the multiple degrees of freedom available that the patient can use to improperly compensate for hemiplegia. Recovery after hemiplegia begins in the proximal portion of the arm. The proximal shoulder regains function before the distal forearm and hand. Many people with stroke never regain complete function, and will elevate the scapula and abduct the shoulder instead of extending the shoulder and elbow. While this may contribute to the end goal of reaching, the shoulder flexors/elbow extensors, the biceps brachii, do not activate properly. The proper activation of the biceps brachii is an important therapeutic goal [47].

In addition to physiological goals, patients benefit from behavioral therapy goals. Many patients learn non-use, in which the patient does not use the paretic arm because

the patient has acclimated to the idea that it is no longer functional [62]. Reminding the patient that their arm can be used is a valuable retraining of behavior, as well as a physiological goal. This concept can be put into practice with mental imagery.

1.2.4 Imagery as a Rehabilitative Task

Mental practice (MP) involves cognitive rehearsal by way of imagining the performance of a motor task without actual movement. Page 2001 [87] posits that MP may work, because similar motor pathways are activated both in cortical [30] and supraspinal motor pathways as would be activated with actual performance of the task. Page also found that motor learning has been reported with MP of a particular skill combined with physical practice of the same skill [5; 125]. Programs combining mental and physical practice have been shown to be more facilitative of motor skill relearning in rehabilitative settings, including for CVA patients, than physical practice alone [86; 118]. The study data used in this research were collected to test the effect of MP on function in chronic stroke subjects.

Page's case study in subacute stroke found reductions in impairment [85]. Page indicated that a viable hypothesis for this improvement may be concurrent activity occurring in the musculature, and the reactivation of appropriate neuromotor pathways. In order to determine whether improvements occurred in the neuromuscular portion of the motion in addition to known cortical changes resulting from physical practice, Page set out to study improvements in reach tasks. One variable for measuring improvements was sEMG of the affected arm. For comparison, Page used the less-affected arm as a close to normal standard. In order to eliminate concomitants from lifting objects or abduction due to maneuvering towards a target on a table, Gail Forrest developed the

simple pointing task that was used here. The task required the patient to point straight ahead at a comfortable pace. The patient started from rest, with the lower arm resting on the lap.

In response to Page's desire to asses the extra-cortical motor pathways involved in reaching tasks, a method was developed for automatically finding onset that has better clinical value than standard ones that [21; 54; 121] all found to be as reliable as manual detection. These standard methods will be addressed in Section 5.5.

1.2.5 The Less-Affected Arm of Stroke Patients

There is evidence that the 'unimpaired' side of stroke patients also suffer motor deficits [19; 29]. However, the deficits are due to cortical control difficulties, not afferent motor control patterns [21; 33; 34; 83; 111; 112].

The delay occurring during stages one to two of a reaching task performed by the paretic arm (defined in Sub-section 1.2.4) has been found to be a result only of the CVA, not of damage in post-spinocortical pathways. Even on the unaffected side, recent studies have shown that the less-affected arm of a hemiplegic patient still suffers from motor delays, despite common acceptance of the those limbs being "unaffected" [21; 111]. However, these delays are due to poor motor control in the brain, not neuromotor pathways in the motor unit, which is recorded in EMG. For instance, in response to tendon taps, Thilmann [111] identified increased amplitude of excitation in the less affected biceps brachii that was similar to that of normals, with no unusual change in the amount of latency in comparison to normals, even though the *length* of the response was shortened.

Low dexterity in CVA patients is characterized by poor motor activation, but is mostly due to the superior cortical control maps being damaged, not the spinal output. In addition, contrary to commonly accepted relationships between latency and speed, Canning [19] found that there was no significant difference between magnitude or speed of sEMG activation of the biceps brachii between high dexterity stroke subjects and normal controls, when the movement required no force generation, such as needed to lift an object. Furthermore, speed did not affect the outcome. But with lower dexterity subjects, the muscle activated more quickly. The excessive activation (manifested as amplitude in the sEMG) present in low-dexterity subjects manifested as co-contraction may reflect loss of skilful muscle coordination, more than a loss of the ability to innervate the muscle compensated with more intense muscle activation. Similarly, in the study done here, patients with slower endpoint speeds (speeds measured at the end of the limb) may have shorter latencies. The low-dexterity subjects in Canning's study were able to activate muscles fast enough to maintain speed as compared to controls. Like Thilmann, Canning [19] found that this meant that the problem of poor timing was due to the inability to selectively choose muscles, as opposed to reduced speed of muscle activation.

In normal subjects, the electromechanical latency measured is the time from reception of the cortical command into the motor neuron pool before tension output at the limb. Electrical response and resulting muscle recruitment are present before there is enough input to induce a mechanical response (Hudgins 1993 in [38]). There has been a wide range of reports of recorded latencies between the onset of EMG activity, and the point when enough EMG activity is present to produce motion.

This latency period has been reported to be approximately 40 – 110 ms [59] during cyclic arm cranking. Flanders 1991 reported that the variability in latency of the biceps brachii is the greatest of other upper arm muscles, and was approximately 100 ms before the motion. The variability is due to the biceps brachii crossing both the shoulder and elbow and acting as a flexor at each joint. The total delay in Dewald's [34] study, which included the transmission through the reflex arc and excluded motor processing, was 80-100 ms for all muscles. EMG onset latencies for the biceps brachii in healthy subjects were about 87 ms after stimulus onset. Dewald found torque onset latencies of elbow flexion withdrawal in response to stimulus to be 108.9 ms (SD 15.2 ms) in the unimpaired arm. Shoulder flexion latencies were 99.3 (9.1). Torque onset latency was defined as the latency between EMG onset to torque output, as a response to a stimulus applied to the reflex arc at the finger. Applying stimulus to the reflex arc bypassed potential motor processing delays as well as inhibitory activation of the gamma neuron. Latencies of the less-affected arm mirrored normals, and were dichotomized into short and long latencies. Short EMG latencies were 40-90 ms, and long, were 100-120 ms after stimulation, in normals. Dewald explains that there are discrepancies in these ranges due to experimental techniques, including the applied impulse type.

In wrist flexion of the less affected arm, Chae [21] found initiation to occur an average of 263 ms (ranging between 140 and 620 ms) after audio cue. These delays included both motor processing and efferent mechanism, not the time after efferent mechanisms between sEMG onset and tension output, studied here. Chae controlled for lesions that would contribute to delay in the initiation of contraction, because those should be intact for non-caudal CVA. Hammond [51] found recruitment times recorded

with indwelling wires in the less affected arm of the agonist forearm to be slow compared to normal controls, with the less-affected recruitment times to be 466 ms, and controls, 261 ms. Hammond found that longer (or slower) tasks have longer recruitment time. Hammond's measurements included the time from first MUAP, the beginning of MU recruitment, and then motion. This is not in conflict with Canning, who stated that slow speeds may have shorter latencies, because Canning recorded sEMG activity, which records information only after muscle MUs begin to respond. Once MUs respond, the ramp from beginning of activation to force output may be just as fast in people with low skill. This is the same latency that was recorded with sEMG here [19]. Latencies of 87 or 90 – 120 ms were considered standard for this research.

Dewald and Wilson [122] both found that the motor flexion reflex pathways in the less-affected side matched those of control subjects. This was confirmed in earlier studies by Chae [21] and Trombly [112]. Lack of cortical motor processing was not overcompensated for with overactive reflexes in the less-affected arm [122]. Dewald's study complemented this by finding that the error was not in the spindle afferent flexor response either. It was not in the muscle spindle that motor disturbances were created. Because torque onset after stimulation to the flexor reflex arc is similar to normal in the less-affected arms, sEMG data from the less affected arm of stroke patients were sufficient as models for the development of an onset detection method.

CHAPTER 2

WAVELETS

2.1 Introduction: What is Wavelet Analysis?

It is generally agreed that movement onset is reflected in the sEMG signal. Physiologically, DeLuca [31] defines activity as enough recruitment and increased firing rate resulting from more recruited motor units, to induce tension output. The manifestation of activation in a sEMG signal has been loosely defined as the point in the EMG signal at which the signal deviates from baseline. Despite years of research, the amount of deviation required, and from what baseline, is still under discussion. Furthermore, Winter [123] indicated that the frequency content of the EMG relates to generated tension, meaning that a simple amplitude change will not sufficiently reflect either onset or magnitude of the tension generated. Because of the combination of influences, more suitable detection methods have been an ongoing pursuit. This is apparent in the variety of processing methods that have been developed over the years, one of the newest, being wavelet analysis.

Wavelets can be used to identify frequency changes occurring at a particular point in time. Event detection in wavelets finds local extrema or inflection points in the generated coefficients of a signal [4; 43; 55; 77; 90; 98].

Wavelet analysis can be used as a form of filtering [43]. It is applied as a vector defined by a known number of elements. A wavelet function is an orthogonal set of waveforms of finite length and an average value of zero. The inner product (convolution) of the wavelet vector and sections of the signal s(t) results in a series of coefficients that describe the quality of the fit of the wavelet vector to the signal, at each point in time.

The coefficients of the wavelet and any companion functions are used for designing the filters applied to raw signals. The wavelet shape is designed to correlate with rapid fluctuations occurring at the center of a sEMG burst.

2.2 A Brief History of Wavelets

Although the basis of wavelet theory is in the Fourier Transform, wavelet theory has only been in a functional form for about 15 years. The roots of the theory are from the 70's, beginning with J. Morlet's alternative to the short-time Fourier analysis [98]. Morlet's goal was to gain temporal resolution for high frequency transients, an essential component of biosignals, and good frequency resolution for low frequency components [98]. Both of these cannot be accomplished simultaneously in Fourier transforms, due to Heisenberg's uncertainty principle. Instead of using an infinitely long wave, Morlet took a precut (windowed with a smooth window) cosine wave, of defined length, and used that as the transform function [113].

The first time a 'wavelet' was discussed was in 1909, by Alfred Haar. His was the first recognition that isolation of signals can be done with scales, or functions other than windowed infinite signals. Haar defined a basic principle using a translated and dilated step function. He found that he could generate an orthonormal basis with this function, essential for vector manipulations.

The ability to repeatedly halve the required number of data points comes from an interpolation theory developed in 1935 by Whittaker [109]. It is that ability which allowed the efficient application of wavelets and ease of use of coefficients.

Advances occurred in 1980 when Stromberg found another linear function besides the step wave that also generated orthonormal bases, and gave better approximations. At the same time, Meyer, while trying to disprove that these waveforms were regular (a necessary condition for defining the smoothness of the wavelets), repeatedly found himself developing families of wavelets that were continuously differentiable.

In 1986, Yves Meyer and Stéphane Mallat were responsible for the application algorithms and the formulation of multiresolution theory [98]. Ingrid Daubechies (the designer of the wavelet used here) invented compactly supported orthonormal wavelets for use in discrete wavelet analysis. Ronald Coifman and Victor Wickerhauser are responsible for some universally accepted wavelet methods [80]. Most significantly, wavelet theory has been developed as a result of dialogue between scientists who have often met by chance [109].

2.3 Significance

Before any analysis of the signal can begin, it is simple to define the sEMG signal as a compilation of individually identifiable patterns and noise [40], such that

$$s(t) = \text{noise}(t) + \text{sEMG}(t) \tag{2.1}$$

The number of MUAPs that contribute to a signal vary at each point in time underneath an electrode pair. A sEMG signal can be described as a transient signal, because the same frequencies are not present throughout the entire signal. Mochimaru [81] explains continuous wavelet transforms as a two-dimensional function (frequency and time dependent) of a one-dimensional signal (time dependent only). Non-stationary

biosignals with transient components change both in time and frequency – they are multivariate [2]. Thus, a multivariate method of approach is optimal.

The basic theory of Heisenberg's uncertainty principle stated that frequency and time cannot both be solved for at once [94; 101].

$$\Delta t \Delta f \geq \tfrac{1}{4}\pi \tag{2.2}$$

where Δt and Δf are inversely proportional resolutions. Equation 2.2 states that the time-bandwidth product must be greater than $\frac{1}{4}\pi$; neither will ever be zero [100]. Unless two frequencies are greater than Δf apart, or Δt apart in time, they cannot be discriminated. Thus, if all points in a signal are greater than Δf, but the time window is too short, data is unresolved, and vice versa. Wavelet methods surmount this obstacle with a multi-resolution approach, by changing one variable while the other is considered. A two dimensional range of coefficients is presented, so a picture across both planes can be compiled into temporally relevant information [101].

2.4 Transforms

2.4.1 Review of Fourier Transforms

The Fourier transform (FT) finds frequencies present in an entire signal, independent of time. Fourier theory represents a signal using orthogonal basis functions involving sines and cosines [96]. The coefficients $\Im(j\omega)$ depend on the signal itself and the basis functions. A power spectrum gives the strength of each frequency present in the signal (a sample in Figure 1.8). The spectrum is a histogram of the calculated ω present, against $|\Im(j\omega)|$ (see Equation 2.3). It allows the approximation of which frequencies of the most contributory to the signal, with strong presence represented by high coefficients. This

method describes the match of s(t) to a set of sine and cosine waves of the form $e^{-j\omega t}$. This power spectrum is a standard source of sEMG analysis.

The FT produces a continuous amplitude spectrum with Equation 2.3 [2],

$$\Im(j\omega) = \int s(t)e^{-j\omega t}dt \tag{2.3}$$

where ω is a form of $2\pi f$, in unitless terms of radians per second instead of cycles per second. ($2\pi f = 2\pi/T$, where $f = 1/T$ – the inverse of period).

The coefficients are determined by convolution between the transform function and the original signal, which is the inner product of the two vector arrays, such that

$$\langle s(t), g(t) \rangle = \int s(t) \bullet g^*(t)dt \tag{2.4}$$

where the one dimensional $g^*(t)$ is $e^{-j\omega t}$ in a transposed array, and s(t) is the signal. The convolving function in Fourier Transforms is a sine wave basis function of infinite length. The plots of one known sine wave may generate a very simple power spectrum with a peak at just one frequency. However, with more complex signals such as sEMG, the spectrum will not simply identify a few main frequencies, but will be spread out over a range of frequencies with a concentration of higher coefficients in one or two areas (Figure 1.8).

If any transient is present in that window of time on the signal, the frequency contents will be spread all over the power spectrum with no temporal resolution.

In order to adapt Fourier methods to shorter windows of time, and thus possibly isolating transients, the method of Short Time Fourier Transforms (STFT) was created. STFTs are carefully windowed finite segments of the sine and cosine waves, slid across

the signal as in Figure 2.1. The window, represented by g, is usually a Gaussian shaped function [101].

$$STFT(\tau,\omega) = \frac{1}{2\pi}\int s(t)g^*(t-\tau)e^{-j\omega t}dt \qquad (2.5)$$

where τ is an incremented time shift nt_0, and ω an incremented frequency $\omega_0 m$.

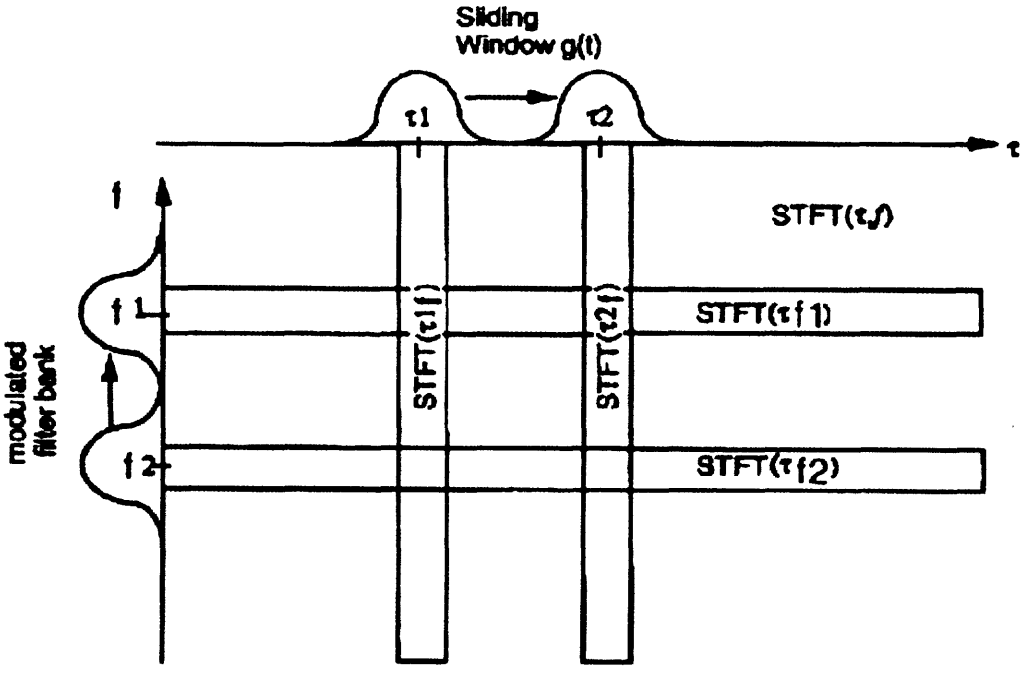

Figure 2.1 Across time, the STFT is a sliding window. Across frequency, it's a modulated filter bank.
(Source: Moshou; Rioul) [82; 101]

The frequency halfway between the low and high end of the power spectrum, the median frequency, is used to represent the content of the signal. A shifting of median frequency is commonly used to identify muscle fatigue. This however, is not useful in small simple motions, such as a single reaching task, or continuous and standard motions, such as gait, which require more temporal resolution, and where slowly changing frequency content is not of interest.

The windowing in time causes a lack of frequency resolution as Heisenberg's uncertainty principle demonstrated. Loss of resolution is due to the fewer available samples in a confined time frame, and to the addition of oscillations in the tails of the Gaussian window that change the characteristics of the transform itself when it is applied over increasing frequencies, evident in Figure 2.2 c. It is difficult to locate both transients and long duration events, simply because a single window size is chosen in STFT. Signals obtained during dynamic conditions do not have features occurring on the same scale throughout, and thus, STFT is not useful [89].

The difference between the translation in two dimensions of time and frequency for STFT and WT, is the method of control of the transform functions, as shown in Figure 2.2 b and d. [2; 101]. In WT, faster frequencies are identified with shorter waveforms, maintaining a low time-bandwidth product, as opposed to STFT, where the waveform size stays the same, increasing the time-bandwidth product. When the length of the transform for identifying high frequencies is long, short duration transients will not generate coefficients large enough to be detected above baseline noise.

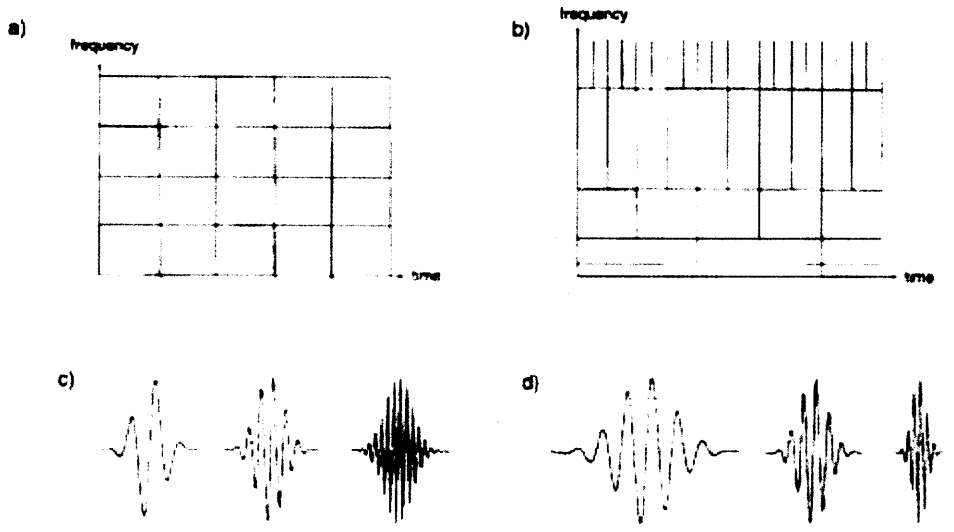

Figure 2.2 a) is the incremental and uniform sliding in two planes to form STFT. The transform function is in c) which is a Gaussian window of a sine wave filled in and kept the same size. b) is the time-frequency tiling for WT (in particular, DWT) that demonstrates how both time and frequency change, while maintaining area, using transform function d), which is not padded with extra oscillations.
(Source: Rioul) [101]

2.5 Wavelets as Transforms

Compared to STFTs, WT is better suited for finding high frequency transients, or long duration, slowly varying signals [96]. The FT and STFT were only able to extract frequencies present in a block of time, not very helpful for short duration and high frequency changes such as transients or non-periodic waves. The STFT is most useful for continuously present frequencies, but not for short duration pulses. Biosignals contain a combination of impulse-like events (spikes and transients) and more diffuse oscillations (murmurs, repeatedly firing neurons), all valuable information.

The wavelet method modulates both time and frequency. The significant difference between WT and STFT is the transform function convolved with s(t) in each of the two methods. Both transforms are linear. However, the STFT uses infinitely long

sine waves cropped with secondary functions, and the wavelet method uses bounded finite known waveforms which reduce the effect of the windowing, because the function is not cropped and altered at its edges.

Application of the WT transform is essentially a correlation analysis between the input signal s(t) and the basis function of the mother wavelet $\psi(t)$, in various forms [65]. Wavelet matching encompasses two variables – width or scale (**a**) of the wavelet (ψ), and the position or shift (**b**) in time of the original, or mother, wavelet.

Instead of using the $e^{-j\omega t}$ windowed with g and time shifted with $(t-\tau)$ in Equation 2.6, the wavelet ψ basis function is stretched with **a** and time shifted with **b**, then convolved with the original signal in the same format used for STFT [28].

$$C(a,b) = \frac{1}{\sqrt{|a|}} \int s(t) \psi^* \left(\frac{t-b}{a} \right) dt \qquad (2.6)$$

where C(a,b) are the resulting coefficients describing the quality of the match. Scale a is the size of the wavelet, and b is the shift in time (similar to τ in Equation 2.3). The wavelet integral measures variation in the signal, in the neighborhood of **b**, rather than the absolute t. This is important for detecting transients, because local variations are the component of interest.

The coefficients can be thought of as correlation coefficients, and it is here that this method shows practical value. The local extrema of C are inflection points of $s(t) * \psi$. These peaks can be used to find initiations of signals [66; 73; 77]. A wavelet will match well when there is a change in frequency, because it is a finite wave, itself changing from zero to a known frequency.

In WT, a coefficient C describes the fit of a finite wavelet to the same size segment of the signal at each point in time. Whenever the wavelet matches more closely, a higher coefficient is achieved. The wavelet is designed with at least one oscillation, so that high coefficients correspond with oscillations in the signal. A wavelet with one oscillation will match with many locations along a signal, and a wavelet with more oscillations will give higher coefficients at locations where there are several oscillations in the signal at that time.

The structure of wavelet theory allows for the individual design of the mother wavelet to suit desired detection, provided the basis function fulfills a certain set of properties [115].

2.6 Wavelet Properties

There are two basic properties of the wavelet function. One, that it oscillates due to 1) the admissibility condition, and second, that the wavelet must be non-zero somewhere, defined by the 2) regularity condition.

2.6.1 Admissibility Condition

This condition states that the wavelet oscillates about a mean of zero [28].

$$\int \psi(t)dt = 0 \tag{2.7}$$

There are no coefficients at the 0 frequency of the Fourier transform of the wavelet (Ψ). This means that the spectrum of the wavelet in the frequency plane must have a bandpass quality, i.e. a limited range of power, implying that:

$$\left|\Psi(\omega)\right|^2\Big|_{\omega=0} = 0 \tag{2.8}$$

Frequencies at or near 0 would be baseline shifts, or very long slow waves that near infinity in wavelength. If there are no infinite frequencies present, then the wavelet vanishes at some definite point in time, i.e. *localized* in the time domain and *compactly supported* in the frequency domain. There is no danger of windowing effects as a result of limiting the length of the wavelet. The function's amplitude oscillates evenly, in a wave, about 0 [28], with no baseline shift, (Top figure in Figure 2.3) satisfying the admissibility condition. The area is therefore also 0 [98], evident in Equation 2.7. If the function satisfies Equation 2.7, then you can decompose and reconstruct a finite signal $\psi(t)$ from the coefficients without loss of information.

Note that the properties of localization in time and compact support in frequency cannot both be exact, which would contradict Heisenberg's uncertainty principle. For this reason, WT has good time and poorer frequency resolution at high frequencies (low scales), and good frequency and poorer time resolution at lower frequencies (higher scales). Practically, this means that the area of the boxes in Figure 2.2 a and b are all lower bounded by $\frac{1}{4}\pi$ (Equation 2.2).

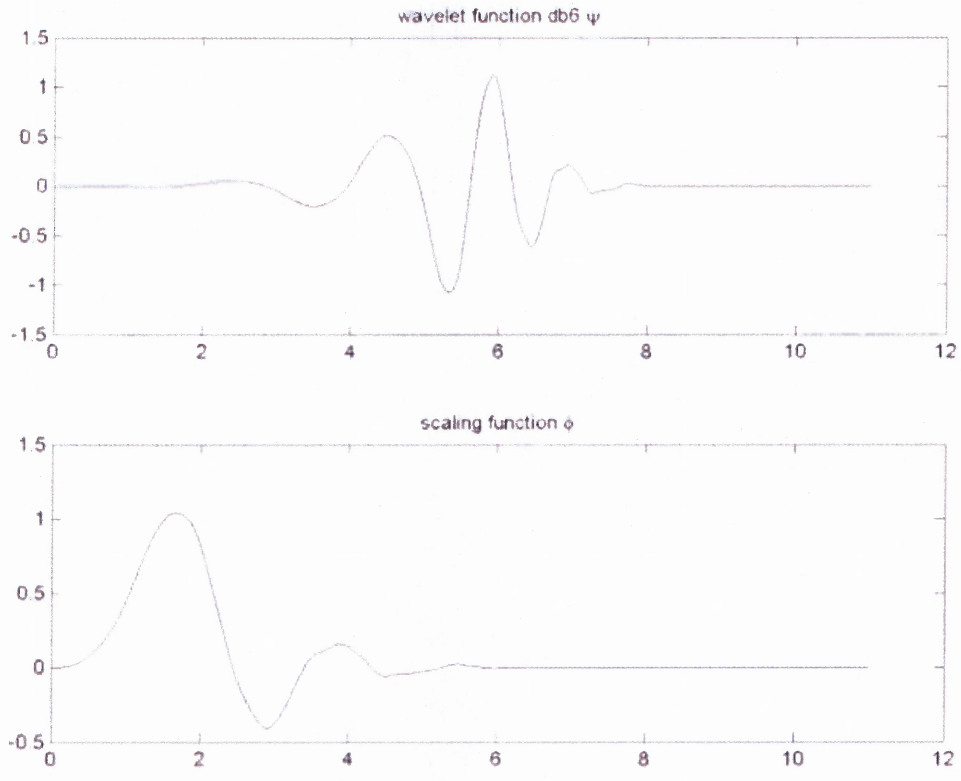

Figure 2.3 Wavelet db6 (top) and accompanying scaling function (bottom).

2.6.2 Regularity Condition

Regularity attempts to keep the size of the time-bandwidth product (Equation 2.2) as low as possible by maintaining concentration in both the time and frequency domains [94]. One way to do this is by providing a way to eliminate low frequency components from high scales. The condition indicates that at very small scales approaching zero, the wavelet must decay quickly to an amplitude of zero. Simply, regularity states that the derivative of wavelet must be zero somewhere. As the scale converges to zero, the coefficients decay [98].

Vanishing moments will be used to describe regularity. The vanishing moments are points p at which the integral across time to the power of p is equal to zero, when the

area of t^p is zero, because from Equation 2.7 of the admissibility condition, the integration of the wavelet must give zero.

$$M_p = \int_0^\infty t^p \psi(t)dt = 0 \tag{2.9}$$

Where $s^p(t)$ is the pth exponent of the signal s(t), and Mp is the moment at that point in time. This can be expanded to a Taylor series, with time shift τ set to 0 [113].

$$C(x,0) = \frac{1}{\sqrt{x}}\left[\sum s^p(0)\int \frac{t_p}{p!}\psi\left(\frac{t}{x}\right)dt + O(n+1)\right] \tag{2.10}$$

Where s is the signal, x is a scaling factor, decreasing throughout the series, and ***O*** represents the remainder of the series. Replacing Equation 2.9 into 2.10, the series is expanded into

$$C(x,0) = \frac{1}{\sqrt{x}}\left[s(0)M_0x + \frac{(s^1(0)M_1x^2)}{1!} + \frac{(s^2(0)M_2x^3)}{2!} + \ldots + \frac{(s^n(0)M_nx^{n+1})}{n!} + O(x^{n+2})\right] \tag{2.11}$$

Quick decay is accomplished by ensuring that several terms in the expansion drop out, resulting in the decay of coefficients at those fine scales. All terms in Equation 2.11 with vanishing moments (M_p) begin to drop out [94; 98; 114; 115]. All that remains is $O(x^{n+2})$. The resulting $C(x, \tau)$ coefficients of the transform will therefore also decay at a rate x^{n+2}, and a higher n will decay faster. The more regular a wavelet function, the faster the convergence [101]. Each factor in Equation 2.11 represents a different "frequency" of the wavelet.

For p = 0,...n, the wavelet has n+1 vanishing moments, or a regularity of n+1. Polynomials of degree n are suppressed. Each vanishing moment limits low frequency

(high scale) components in the wavelet, because terms that correspond to low frequencies will drop out of the wavelet transform.

2.7 Application of the Wavelet Transform

Previously, the wavelet method was described as the convolution of a sections of a signal s(t) with a wavelet. Section 2.4 demonstrated how this method involves stretching and sliding the wavelet across the original signal with frequency parameter (scale) **a**, and the time parameter (translation) **b**.

When **a** is large, the basis function becomes a stretched version of the original mother wavelet, which is useful for analysis of low-frequency components of the signal (s(t)). Similarly, a small **a** will contract the basis function, and make it useful for isolating high frequency activity. In general, however, the shift and stretch of the wavelet means that a large **a** corresponds to low frequencies, and a low **a** to high frequencies. There are three modes of applying the wavelet transform [28; 101].

- Continuous – continuous time and scale parameters, defined by Equation 2.6. This is mostly a theoretical application.
- Discretized Continuous Wavelet Transformation (CWT) – used in wavelet expansions; It applies a digitized and thorough transform, to the point of redundancy. This one will be used for transient and event detection.
- Discrete Wavelet Transformation (DWT) – Uses filters based on the coefficients of the wavelet and an accompanying scaling function. The filters are use for the decomposition and reconstruction of the original signal. During reconstruction, various methods can be applied to eliminate coefficients in each level, thus filtering. This will be used for the separation of salient and non-salient signal content.

2.7.1 Continuous Wavelet Transformation

Theoretically, the continuous method means convolving and integrating across all **a** and **b**. Because the WT of a signal with one dimension t generates a transform with two dimensions ω and τ (represented by **a** and **b**) (Equation 2.6). This means that because the wavelet applies all dilations at every sample, each level in CWT analysis maintains the same number of coefficients in the array, and there is high redundancy in CWT. For this reason, CWT is sometimes referred to as 'over complete'.

Dilation **a** is not necessarily positive, nor is it in integer increments, enabling a perfect reconstruction of s.

$$S = \frac{1}{C_\psi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{da\, db}{a^2} < s, \psi_{a,b} > \psi_{a,b} \tag{2.12a}$$

where

$$\psi(t)_{a,b} = |a|^{-1/2} \psi\left(\frac{t-b}{a}\right) \tag{2.12b}$$

In order to apply this concept to digitized signals, a discretized form of the continuous WT (CWT) and the more efficient discrete WT (DWT) was developed. Both CWT and DWT were used in the method developed here, but for different purposes.

2.7.2 Discretized Continuous Wavelet Transformation

Discrete methods are not simply sampling of the continuous method. Discretization is a numerical method meant to accomplish what was theorized by the continuous method [109], necessary for digital signals. This method finds coefficients for every integer increment of scale **a** beginning with 1, and at every time sample **b**, beginning from t=0. Successive scales are not downsampled in any way, so the coefficients arrays cannot be

summed or reconstructed. This method is often called overcomplete, because of the redundancy.

The discretized form of Equation 2.6, with no assumptions about **a** or **b**, is

$$CWT\ (a,b) = a_0^{-m/2} \int s(t)\psi\left(a_0^{-m}(t - nb_0a_0^m)\right)dt \tag{2.13a}$$

Equation 2.13a reduces to

$$CWT\ (a,b) = a_0^{-m/2} \int s(t)\psi\left(a_0^{-m}t - nb_0\right)dt \tag{2.13b}$$

The time shift parameter is $\mathbf{nb_0}$ – an incremented $\mathbf{b_0}$. Parameter **a** is dilated in integer powers of a fixed initial width a_0, such that $a = a_0^m$. The width of the mother wavelet $a_0 = 2$ for Daubechies wavelets, to ensuring orthogonality. The width of the wavelet ψ, represented by $a_0^{-m}\,t$, is proportional to a_0^m. Time shift parameter **b**, can be discretized similarly, so that $b = nb_0a_0^m$. Integers m are all real, and n are all positive real integers [28].

Constable [27] concludes that the CWT is useful for event detection, because the coefficient arrays are not downsampled at each successive level, and can therefore be associated with real events, such as onset, in the original signal.

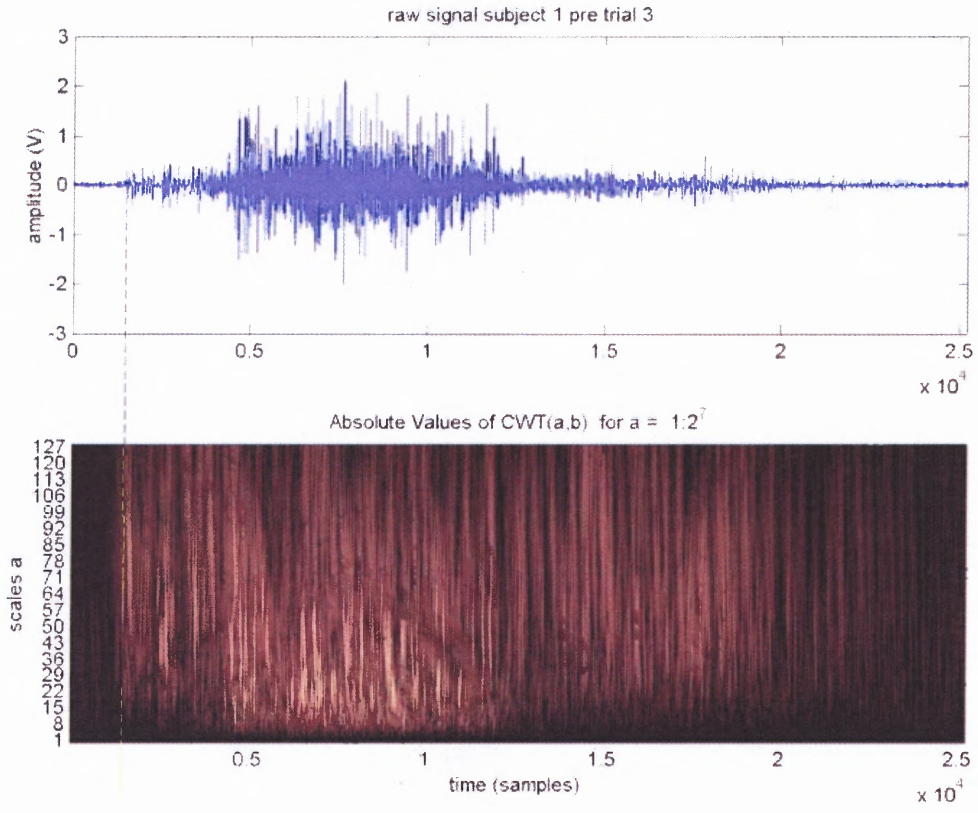

Figure 2.4 The top plot is the raw sEMG for which the CWT coefficients have been calculated in the bottom plot. The time base is in samples, because the translation of the wavelet across time is done for every sample. The dotted line is the point detected as activation by the method. That point is a peak in the wavelet coefficients, identifiable by brighter regions. The integer step of parameter m for base 2 gives the ordinate axis a logarithmic scale.

2.7.3 Discrete Wavelet Transformation

For more efficient filtering than the overcomplete method, the DWT, also known as 'filter bank' expansion, tiles the application as in Figure 2.2b. The method of filtering accounts for redundancy, and allows successive downsampling which reduces time of calculation [9; 98]. Resolution at lower frequency scales is reduced as well [27]. DWT is valuable for filtering out or reducing the weight of non-salient bandwidths. The filtering capability was used to extract a band of noise for SNR calculations, before the CWT was

applied to the filtered signal. The CWT slides the wavelet across the time domain in single unit integers. In contrast, because of the dyadic nature of the filtering (halve, then halve then halve etc), the plane of time in DWT is in increasingly smaller widths. Figure 2.2b shows the variable widths in translation [109].

Compression in both time and frequency domains requires two orthogonal functions. In DWT methods, the coefficients of those functions are used to build a set of four filters. Because each set of high and low filters, one pair for deconstruction, another for reconstruction, are inverses of each other, the arrangement is called a quadrature mirror filter bank.

The scaling function accompanying db6 shown in Figure 2.3b is used for the low pass filter to get approximation coefficients (A). The wavelet function describes the coefficients for the high pass filters to get detail coefficients (D)[101]. Both of these functions must be orthogonal.

Orthogonal functions [68] are functions g that

$$(g_m, g_n) = \int_a^b g_m(x) g_n(x) dx = 0 \tag{2.14}$$

and $m \neq n$, on the interval $a \leq x \leq b$. This is equivalent to Equation 2.4 being equal to zero.

A set of real valued functions are called an orthogonal set of function on the interval, if these function ($g_1(x)$, $g_2(x)$, $g_3(x)$) are defined on that interval $a \leq x \leq b$, and if all the integrals exist and are zero for all pairs of distinct functions in the set. An orthonormal set of functions is one where the norm of each function, $||gm|| = 1$. Norm is defined by

$$\|g_m\| = \sqrt{(g_m, g_m)} = \sqrt{\int_a^b g_m{}^2(x)dx} \tag{2.15}$$

Thus, an orthonormal set of functions on the interval $a \le x \le b$ satisfies the relation:

$$(g_m, g_n) = \int_a^b g_m(x)g_n(x)dx \begin{cases} = 0 \text{ when } m \neq n \\ = 1 \text{ when } m = n \end{cases} \tag{2.16}$$

Dividing each function by its norm on the interval under consideration forms an orthonormal set from an orthogonal set. Sine and cosine waves, of the form 1, cos x, sin x, cos2x, sin2x, form an orthogonal set. If one of the g functions can be represented in terms of g_j by a convergent series as a "generalized Fourier series" $f(x)$,

$$f(x) = \sum_{n=1}^{\infty} c_n g_n(x) = c_1 g_1(x) + c_2 g_2(x) \ldots + c_n g_n(x) \tag{2.17}$$

then the coefficients c_1, c_2, etc, are the Fourier constants f_0 with respect to that orthogonal set of functions. In fact, the ability to represent a set of orthogonal functions by convergent series, allows the use of coefficients in wavelet transforms. The g are orthogonal functions, similar to the sines and cosines in the Fourier series. Each factor represents a different frequency. In fact, Equation 2.11 has the same property, which shows how WT coefficients can represent a signal, like coefficients from a FT. The constants **c** can be determined easily, using the same function [68]. All function coefficients are independent from each other. These constants are the coefficients used in WT.

These coefficients characterize the original signal thus:

$$(f, g_m) = c_m \| g_m \|^2 \tag{2.18a}$$

which can be reconfigured to

$$c_m = \frac{(f, g_m)}{\|g_m\|^2} \tag{2.18b}$$

The easy manipulation of coefficients becomes useful and necessary during filter design and DWT.

The basic idea of DWT filter application, is a sequence of halved and embedded approximation spaces (Figure 2.5) [97]. W.H. Press has developed an algorithm for discrete analysis using Daubechies wavelets that is optimal for digital signals, and is similar the method adapted by Matlab (Press in [27]), which was the programming environment used for development in this research. The number of bands, or levels (m_{max}), is dependant on the length of the data vector, or sampling frequency, similar to Fourier transforms. Each band of frequencies (each level) is half the sampled points of the previous, as well as half the mean frequency. The frequency bandwidth of the first level is $\frac{1}{2}F_{max}$ to F_{max}, where F_{max} is the maximum frequency present in the signal. According to the Nyquist theorem, a $2*F_{max}$ sampling frequency is required in any signal to prevent aliasing, just as needed for FT. Data sampled over a bandwidth of 10-1200 Hz must be sampled at least at 2400 Hz. The mean bandwidth present in the top level is half that: 600 – 1200. Each level (of one set of approximation coefficients A and one set of detail coefficients D) is generated when the original signal is passed through a set of filters, arranged in groups, or banks. The signal described at each level is now a series of independent banks of values that can be examined separately.

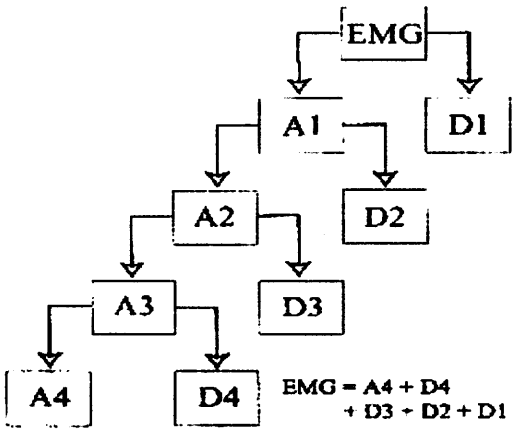

Figure 2.5 Decomposition tree. Each level is a decomposition of the previous level. To recreate the original signal, the last level of approximations coefficients can be added to all the levels of detail coefficients, after each band of coefficients has been reconstructed. (Source: Flanders) [43]

The halved length arrays of coefficients at each successive level still efficiently represent the original signal, thus allowing the summation of different banks of detail coefficients, without increasing the length of data after reconstruction. Because the coefficient arrays of orthogonal functions are so easily downsampled, orthogonal wavelets are required for DWT [101].

First, to deconstruct the signal, the original signal is high pass and low pass filtered with filters of which the coefficients are defined by the ψ and ϕ functions. This filtering gives two sets of coefficients, each of the same length as the original signal. These coefficient arrays are downsampled to half their length. If the coefficients were to be reconstructed without downsampling, the resulting signal would be twice as long. This is repeated for the number of iterations specified by maximum level m_{max}, the value of which will be discussed in Section 2.8. Higher levels are of lower frequency content.

For reconstruction, the filters are inversed and applied to upsampled coefficient arrays (Figure 2.7).

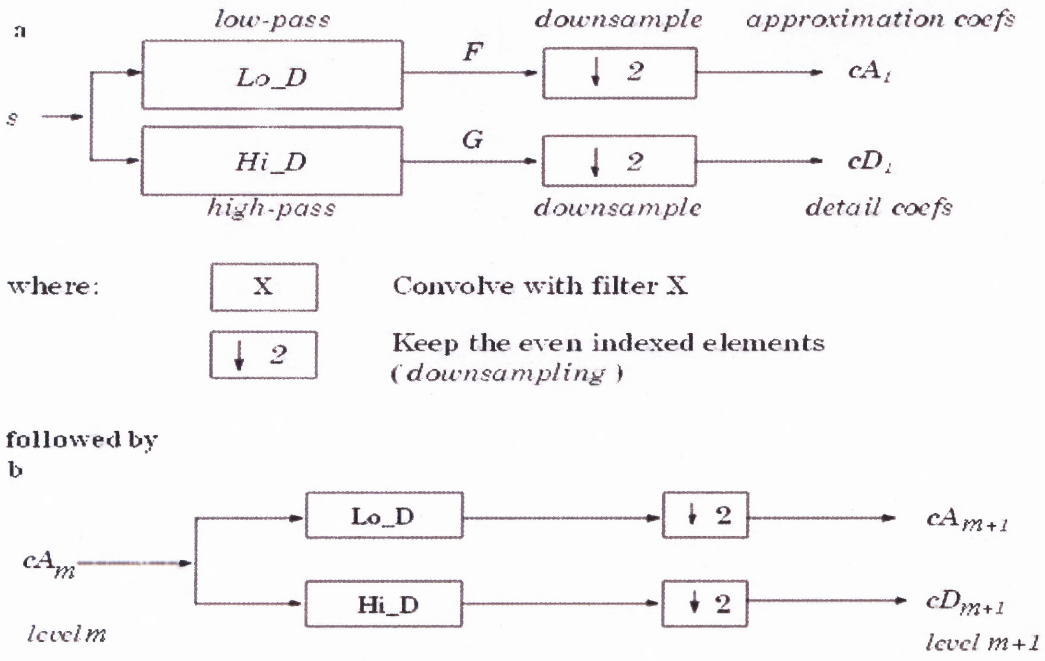

Figure 2.6a,b (a) Lowpass subsampled s, passed through highpass and lowpass filters F and G, from which approximations A (or cA) and D (or cD) are derived. The approximation coefficients cA of each level are then passed through the same deconstruction filters (b).

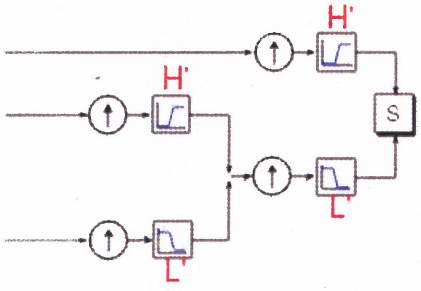

Figure 2.7 Upsampling and reconstruction of the coefficient arrays created in Figure 2.6. The combination arrows indicate that the H' and L' can simply be added to get back to S.

The filter coefficients of both deconstruction and reconstruction are based on coefficients of the wavelet ψ and its associated function ϕ, not standard Fourier based filters. The design of ψ is based on desirable high pass filter coefficients, and produces the details. Conversely, the low pass filters determine the companion function, the scaling function ϕ.

The reconstruction filters are inverses of the deconstruction filters. The filters are inversed by reversing the order of their coefficients then padding with alternating zeros to bring the sampling back up to the original sampling rate. The same is done to the arrays of coefficients representing the signal; they are padded with zeros to double their size, and passed through the inverses of the filters. The upsampled approximation coefficients are convolved with the upsampled ϕ filter coefficient vector, and the detail coefficients with the ψ coefficients. To perfectly reconstruct the original signal s, final reconstructed levels of A and D can be added [80]. A sample set of reconstructed levels can be seen in Figure 2.8. The association between level and approximate frequency (pseudofrequency or pF) on the ordinate is explained in Section 2.8. To denoise the signal, an imperfect reconstruction can be obtained by adding fewer levels [88].

The ability to upsample and inverse the filter coefficients without changing the meaning of the filter, is a property enabled by orthogonality. Daubechies wavelets are orthogonal, thus the construction of conjugate mirror filters with a db6 wavelet is possible [109].

The scaling function ϕ seen in Figure 2.3b is basically an impulse response of a low pass filter. The filter coefficients associated with ϕ that are used for calculating details are the inverse transform coefficients of that response. The coefficients of ϕ are only useful when they are greater than 0, where they are real.

Similar to the wavelet, the scaling function has an admissibility condition that shows that the 0^{th} moment of the scaling function does not vanish [115].

$$\int \phi(t)dt = 1 \tag{2.19}$$

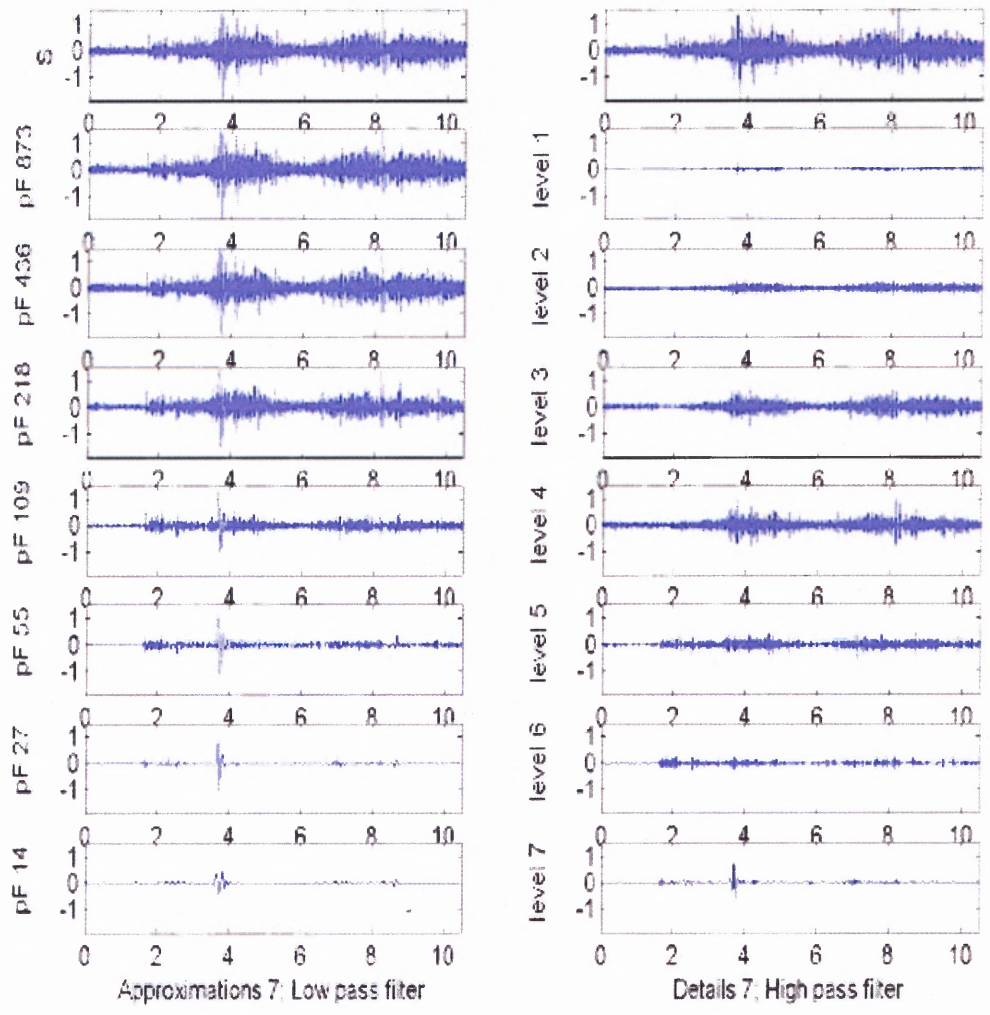

Figure 2.8 On the left side are reconstructed approximation coefficients, and the details are on the right. Each band has an approximate mean frequency, a psuedofrequency, defined on the far left as pF, corresponding to each level.

Prior to the reconstruction in the levels in Figure 2.8, various filtering techniques can be applied to just select regions of frequency within the whole signal by way of thresholds, or by weighting each level of coefficients. In simulated signals, noise is represented with white noise [77]. Similarly, assuming the ideal condition that the noise present in real sEMG is white noise, that noise would be of similar weight in each band. Therefore, looking in bands of sEMG where there is little to no sEMG content for noise

can provide an estimate of noise throughout the signal [41; 114; 115]. Bands representing frequencies over 400 Hz contain such noise, and were used for signal-to-noise ratio calculations. The highest levels of detail, levels one and two on the right side of Figure 2.8, are visibly uniform with noise that approximates white noise. Similarly, in frequency bands lower than 15Hz, such as present in the approximations and the high order detail, most of the content is slow noise or movement artifact, and can be omitted during reconstruction [81]. In the data studied here, hard thresholds were applied by weighting to zero the coefficients found in non-salient bandwidths of information. Since a perfect reconstruction was not the desirable outcome, a filtered signal (s') that was to be passed through CWT analysis was created by summing only bands (generated with DWT methods) containing frequencies of interest.

The filtered signal s' can be reconstructed just from details on the right side of Figure 2.8. The reconstruction of discrete coefficients of the signal, can be done with a discretized version of Equation 2.12 [101]

$$s(t) = \sum_{m \in Z} \sum_{n \in Z} c(m,n) \psi_{m,n}(t) \tag{2.20}$$

2.8 Relationship of Scale to Frequency

The relationship of scale to frequency in continuous and discrete methods is from Equation 2.19a. In the CWT, the change in scale is incremental in units of $a_m = a_{m-1}+1$. When the maximum m = 7, scales **a** range from 1:64. Increments of time relate to the sampling frequency. If s(t) was sampled at 2400 Hz, one time increment $\tau = 1/2400$ [28].

In DWT, scale **a** is dyadic, changing in power increments of one, to base 2. Each scale $\mathbf{a} = \mathbf{a_0}^m$, or, $\mathbf{a} = 2^m$, for m = 1:7. Integer increments are used as the power to which a base ($\mathbf{a_0}$) is raised, as seen in Equation 2.20a.

The length of the array describing the wavelet depends on the sampling frequency of the analyzed signal. Data is sampled at 2400 Hz, thus, the actual length of the wavelet to be convoluted is changed to maintain it's central frequency at that sampling rate.

$$pFa = \frac{Fc}{a\Delta t} \tag{2.21}$$

The approximate center frequency of the mother wavelet is Fc, and pFa is the pseudofrequency of each level **a**.

The sampling period is t, the inverse of the sampling frequency (1/2400 for this study). The wavelet db6 has a Fc = 0.73 Hz. This means that if, in Figure 2.3a, the abscissa were to be considered time, approximately 0.73 turns would occur in one second.

With these values, Equation 2.21 becomes

$$pFa = 2400 \times 0.73 / 2^m \tag{2.22}$$

with **a** ranging with m=1:7 over 2^1:2^7 for the DWT and 1:64 in the CWT. The pseudofrequencies on the ordinate scale in Figure 2.7 are calculated in Table 2.1.

Table 2.1 Calculation of Scale (by Level) to Pseudofrequency

Level	Substitute	pFa
1	$2400 \times 0.73/2^1$	876
7	$2400 \times 0.73/2^7$	13.7

These pFa are the center frequencies of bands of sequentially halving frequencies in DWT. The frequency resolution of the first band is the highest band, ranging from ½ Nyquist (Fmax) to Nyquist (2Fmax). The second band is ¼ Nyquist, to ½ Nyquist, thus a log scale. There is a single output vector containing all the coefficients. These coefficients are used for reconstructing the original signal [27].

The ordinate axis on the bottom of Figure 2.4 shows how the scale varies almost continuously between 2^1 and 2^5 by unitary increases (from the bottom to the top) in the CWT. When a scale is small, only small details are analyzed, as in a geographical map. Because each level is the same length as the original signal, the CWT coefficients cannot be used to reconstruct the original signal, but gives very good information for identifying transients and other events.

2.9 Why a Daubechies Wavelet of the 6th order

Matlab (The Mathworks, Natick MA) has developed a software package for the application of standard wavelet forms to any signal type, including any order of Daubechies wavelet. Daubechies was decided upon as a standard wavelet, due to its similarity to the MUAP [17; 43; 69; 81; 101; 103]. Furthermore, setting $\mathbf{a_0}$=2 ensures that Daubechies family is orthogonal, and can thus be used in both DWT for filtering and CWT for event detection.

Standard wavelets applied to EMG (both surface and indwelling) are in the Daubechies family, of varying orders, or number of turn points. Orders of DB2 [43], DB4 [88; 90; 93] and DB6 [56; 94], DB8, and DB16 [88] have all been suggested.

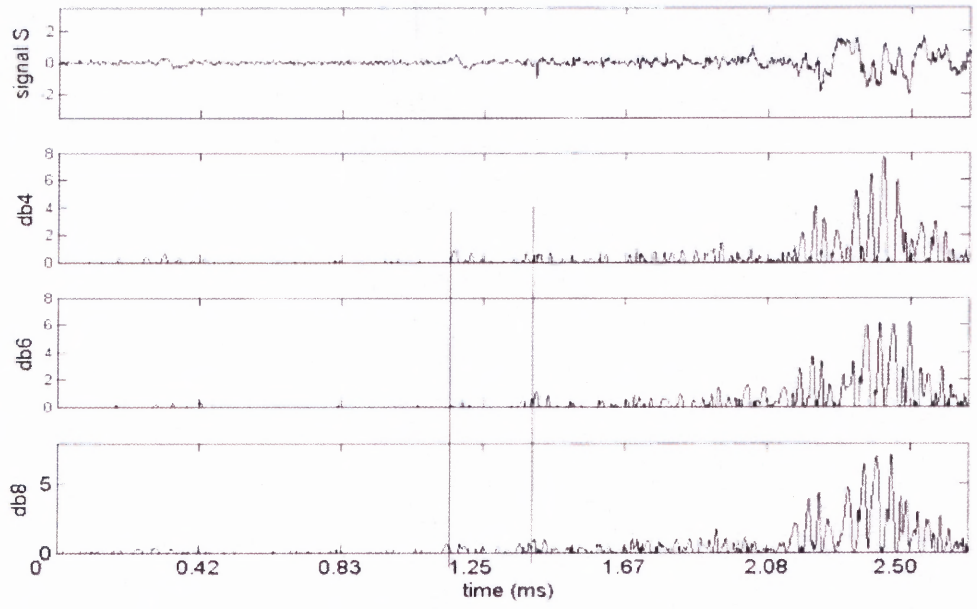

Figure 2.9 Original signal assessed by three wavelets db4, db6 and db8. The black lines indicate the region where onset will occur. Neither db4 nor db8 distinguished artifact from change in frequency as well as db6. Because of the greater regularity of db6, the artifact just before the signal started was not given a falsely high coefficient. Db8 has too many oscillations, thus, the amplitude of the salient portions were not given better significance. Onset time was actually 1336 ms, easily identified with db6.

The smoothness of db6 implies that peakiness in the convolved signal is not going to result in high coefficients, and will be filtered out.

Flanders found db4 more fitting that db2 for biceps brachii activation. Embedded electrodes that read individual MUAPs have been assessed with db2 [17]. Db2 is simple and similar to the MUAP. However, due to MUAP additive properties, the IP related to tension generation in sEMG has a greater density of fluctuations than individual MUAPs, therefore not fit well with db2. Pattichis explained that the db4 is useful for tracking transient components of the MUAP, however, Pattichis used indwelling wires [90]. Lower order Daubechies captures short transients, important in assessing bursts of neuromuscular activity. There exists low frequency noise due to polarization potentials

and motion induced potentials between the skin and electrode. Thus, low order wavelets are inappropriate. MUAP IPs recorded with sEMG have longer durations, thus a higher order Daubechies is better.

Figure 2.9 demonstrates a sample detection with db4, db6 and db8. The detection method will be addressed in Chapter 3. Between the vertical lines is the critical zone where onset should be found. Daubechies wavelets of orders lower than six tended to identify a few bursts, regardless of whether the burst was noise or salient information. The low order db wavelet, of order four, just as easily identified short bursts that do not add up to activity. In sEMG, a group of bursts contribute to tension generation, and unitary bursts, slow single waves, or transients, are undesirable information. Thus, a higher order Daubechies wavelet with more oscillations will identify a set of bursts, representing the IP better. Higher order wavelets than db6 took far more computational time, and also identified too many activations as evident in the bottom plot of Figure 2.9, and similar to lower order wavelets, did not differentiate noise from activity very well.

CHAPTER 3

METHODS

3.1 Subjects

Data were collected under a grant from The Retirement Research Foundation, of whom Dr. Stephen J. Page [85] was the primary investigator. Inclusion criteria included (1) CVA experienced between one and six years prior to study enrollment; (2) no cognitive deficits or gross language deficits, as evidenced by a score of 9 or fewer error points on the Token Test [7], and a score of 70 or more on the Modified Mini Mental Status Examination (MMSE) (3) age between 18 and 95; (4) no excessive spasticity (as measured by a score of "2" or more on the Modified Ashworth Spasticity Scale) or pain (as measured by a score of "4" or more on a Visual Analog Scale) in the more affected limb; (5) completely discharged from all forms of physical rehabilitation; and (6) 10° of active flexion in at least two fingers/thumb and extension back to neutral (anatomical position) in the same two fingers/thumb. The focus of this work was the development of this method rather than the testing of data for statistical robustness, thus time constraints limited this study to two subjects. Separate work of testing multiple subject data for the purpose of statistical assessment of the reliability of the method has since begun.

3.2 Task

Twelve sEMG traces, collected during three trials pre and post training of two chronic stroke subjects, were assessed for improved motor control due to mental practice [86; 87]. The task recorded required the subject to point forward in the saggital plane at a self-selected speed. A saggital plane movement incorporates both shoulder flexors and elbow extensors, both movements engaging of the biceps brachii. Speed was measured by the linear velocity of a marker placed on the second metacarpal joint of the hand. Recording began while the patients arm was still at rest, with the pressure switched placed under the patient's natural resting position in his or her lap.

The onset of motion, as seen in video data, was defined as the frame in which any movement incorporating the shoulder, elbow or trunk was noted. This included trunk shift, which would place a stretch on the biceps brachii. The biceps brachii of the less-affected arm of each subject was studied for onset of EMG activity prior to onset of motion, using the method described below.

3.3 Instrumentation

Twelve traces, collected during three trials collected pre and post imagery training of two chronic stroke subjects, were assessed for improvement in motor control after mental practice [87]. EMG data was sampled at 2400 Hz, twice the maximum frequency content of the signal collected over a bandwidth of 10-1200 Hz. The Dantec silver/silver chloride disposable electrodes were connected to a Gould 6600 Bioamplifier. The amplifier fed into a National Instruments AT-MIO-16X Data acquisition board, and was controlled by a program written in Labview 5.0. The same data was fed into the analog input to the

datastation of the Vicon 370 (Oxford Metrics, Oxford UK) motion capture system, which sampled video data at 60 or 120 Hz. Vicon sampled analog data at 1200 Hz, and was therefore not sufficient for sampling as required by the Nyquist theorem. Synchronization between the two systems was accomplished with a pressure sensitive switch resting on the lap of the subject that passed a continuous sine wave into the analog input of both the National Instruments DAQ and Vicon. When the subject lifted his or her hand, the switch turned off synchronously in both systems, terminating the sine wave.

3.4 Processing and Analysis

Wavelets have mostly been used for identifying recruitment patterns, or the presence of activity in particular phases of activity in signals with higher frequency content. Merlo [77] has adapted the WT for precise identification of initiation of neuromuscular events, and was used as models for this method. Continuous analysis gives a better picture of bands of frequency content [56; 58], and is valuable in feature extraction [98]. Thus, both DWT and CWT were used here.

The predominant frequency band of the biceps brachii is 50-60Hz (16 –20 ms, or 25-166Hz in most muscles which is 6-40 ms [123]) for one sEMG MUAP. There are still frequencies present as low as 10Hz [41]. All levels representing 1-25 Hz, and 200+ Hz, can be assumed to contain noise, and not motion relevant data. Flanders [43] suggested salient levels of 3-7, due to their similarity with MUAP transmission frequency of 20-30 ms, or 15-250 Hz, for surface EMG. Zennaro [126] suggested levels between 2-5 for fine wire EMG, which has a higher median frequency than sEMG. A signal bandlimited with a DWT was then fed into a CWT, eliminating both low and high frequencies, because at

very high scales (and low frequencies), artifacts are more common than information, and with of strong matches at high scales, might distort the maximum coefficient found across all scales.

In most wavelet analysis, the bandwidth associated with short length transients, otherwise known as high frequency noise, can be identified in level 1 of a DWT [43]. Panagiotacopulos (1998) [88] considers the first level of decomposition as the noisy component. On EKG signals, Mochimaru [81] uses hard thresholding on DWT level 7 for denoising, then continues to a CWT for edge detection.

To extract bandwidths of data, or analyze signal content at various bandwidths, DWTs can be applied [27]. DWT methods are frequently used for filtering by hard or soft thresholding, which involves the reduction of the amplitude of the coefficients of any or all bandwidths, prior to reconstruction. Soft thresholds weight each data point below the threshold, to take into account surrounding values. Hard thresholding is more standard for noise removal, and sets all values below threshold to zero. In this case, hard thresholding was applied practically by weighting all coefficients in levels outside salient bandwidths to zero. The signal is not reconstructed to match its original shape, but rather, reconstructed without that bandwidth of noise. After this filtering process, the remaining signal was sent into coefficient extraction by CWT.

One band of noise can represent white noise present in the whole signal [114], because sEMG does include white noise, which has equal power throughout a signal [41]. Therefore, the residual bandwidth was also used as the noise signal for calculating SNR.

Merlo's [77] detection goals most resembled those required for clinical and rehabilitative studies, such as the post MP training reach task tested here. Merlo

generated simulations of single differential surface EMG signals, collected during onset of motion from rest. Analysis was done with a wavelet design based on a Hermite-Rodriguz HR function emulating a MUAP, and the method adapted from Merlo used a standard Daubechies order 6 [93].

For detection, Merlo was used as a model, with a few significant differences. Prefiltering with DWT is not relevant, because the signals are simulated. However, bandlimiting the signal was considered necessary to obtain MUAP durations of physiological reference. After decomposing the simulated and noise signals with CWT, operations were done on the maximum coefficients in both.

Merlo used a parameter γ as a constant multiplier with the maximum coefficient found in the noise signal. The value of the SNR was used for guidance in choosing γ. He found that 1.6 is often optimal, but had no distinct pattern for its applicability. With that in mind, a range of possible γs were applied to the synthesized signals, the result allowing some user control in the choice of which γ to use for constructing the cutoff threshold [40; 77]. In this method, a table is presented to the user where each successive level has a higher threshold, based on γ increasing incrementally by 0.1 (Figure 3.1).

Due to the amplification of the signal information with the transform, the successive subtraction of all information below threshold does not occlude points of onset. In fact, the successive increases in threshold are not critical, as evidenced in Figure 3.1, which shows how the onset point can be found in several increasing thresholds.

The values of γ that Merlo chose were between 1.0 and 2.4, because the SNR used with his method was between 8-12 dB. In contrast, SNR values based on noise in levels

from DWT decomposition ranged much higher, from 12-25. Therefore γ values in this study were much lower than Merlo's, ranging from 0.8 to 1.8.

Another difference was that Merlo's noise signal was an independently selected section of the collected signal. This was not available in the collected data. Furthermore, the ability to use completely self-referential methods is extremely advantageous when dealing with sensitive signals in a clinical environment.

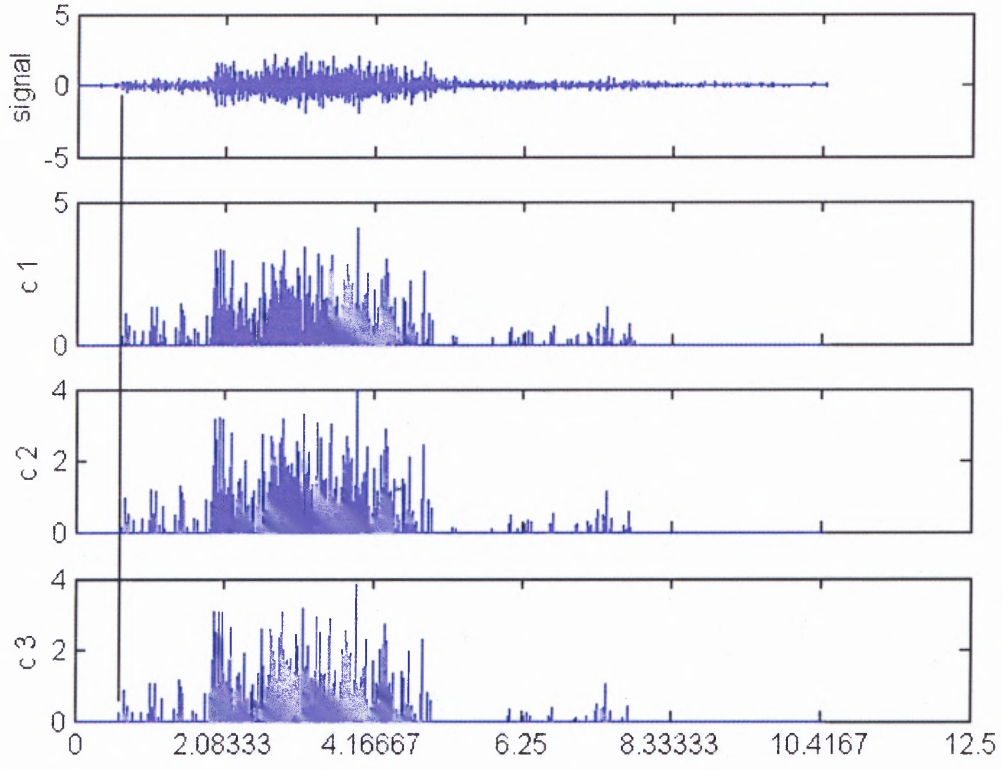

Figure 3.1 Successive iterations of **c** point to increasing values of γ, from 0.8-1.0. Still, the same onset point of 0.62 ms is evident in several thresholded iterations. SNR was 12.25 dB.

Peaks in wavelet coefficients identify changes from inactivity to activity, because high coefficients indicate a good match to the finite shape of the wavelet [55]. The ability to detect changes is because the wavelets are bounded, and any change from zero to

frequency induces a high coefficient. Thus the point in time of the highest peaks across all levels of transformations indicated burst onset [2; 75].

3.5 Coding the Method

This method was applied to all data types used for this study, with the kernel based on a program written by AM Petrock [94]. Figure 3.2 presents a tree of the method of parameter extraction.

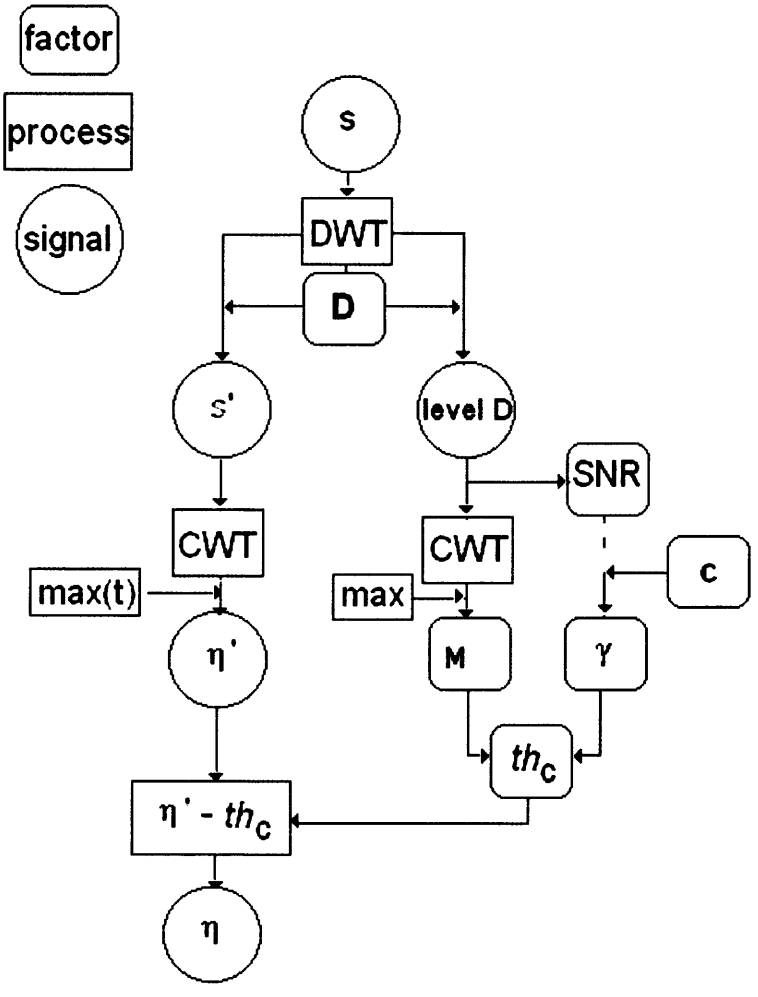

Figure 3.2 Tree identifying the sources of input parameters. The only input parameters **c** and **D,** are based on information in previous steps. The supervised classification is run on the thresholded coefficient array η.

1. First, signal s was sent through DWT and decomposed into approximation and detail coefficient bands. The approximations and details of the signal were reconstructed to represent temporally relevant information present within the signal at several bandwidths.

2. Considering the activity present in each level, any level of high frequency (levels 1 or 2) that had uniform noise and was outside relevant frequency content (anything between 0 – 25 Hz, and greater than 150 Hz) was a valid choice for 'noise level' (**D**). **D**2 was the preferred value, however, if motion related information was present, **D**1 was selected.

3. A bandlimited signal **s'** was built of all remaining levels.

4. CWT was performed on **s'**.

5. Maximum coefficients of each time sample across all scales were extracted and put into an array the same size as **s'** on which activation detection was performed [77].

$$\eta'(t) = \max(\mathrm{CWT}(s',t)) \tag{3.1}$$

6. A hard threshold was built:

   a) The value of the highest coefficient, M, was found across all samples (**b**) and scales (**a**) in the noise band **D** chosen in step 2.

$$\mathbf{M} = \max(\mathrm{CWT}(\mathrm{noise})) \tag{3.2}$$

   b) **M** was multiplied by a factor $\gamma > 1$, defined by Merlo to range up to 2.4, so that the threshold *th* at each iteration **c**

$$th_c = \mathbf{M}\gamma_c \tag{3.3}$$

   The iterative coefficient **c**, ranged from 1:17, identifying a range of possible γ, where $0.8 < \gamma < 2.4$

   c) The choice of γ was loosely based on the signal to noise ratio (SNR) between **s'** and noise
   The SNR is calculated with

$$\varphi^2_s / \varphi^2_n \tag{3.4}$$

   where φ is the mean square value [6], or

$$\mathrm{MSV} = \Sigma x^2 / \ \mathrm{length}\ (x) \tag{3.5}$$

7. Considering the range of *th* constructed in step 6, a table of several versions of η'(t) with any values below th_c set to zero was presented to the user, thus:

$$\begin{aligned} \eta_c(t) &= \eta'(t) - th_c \quad for \quad \{\eta'(t) > th_c\} \\ \eta_c(t) &= 0 \quad for \quad \{\eta'(t) \leq th_c\} \end{aligned} \tag{3.6}$$

8. A $\eta_c(t)$ was chosen, by picking an iteration **c**. Criteria for selected threshold were

    a) To keep the threshold as low as possible.

    b) To use the same **c** for all data

    c) To remove as much noise as possible without losing information. Noise was apparent by spikes unassociated with large clusters of activity. For example, the bottom plot in Figure 3.1, iteration **c** = 3, the bottom trace, shows the same salient activity as the previous traces, but removes more random spiking at the tail of the activity.

9. Peaks of $\eta_c(t)$ were found. A peak signified a change in frequency, thus on or off of the sEMG signal.

10. Even in the most recent methods, supervised classification for a final step is acceptable [126]. Thus, a secondary algorithm was applied for aid in rejecting spurious peaks and identifying groups of activity defined by the peaks found in step nine. Flanders uses a minimum 50 ms distance between peaks to identify separate events in biceps brachii data collected during reach trials. Merlo considers all peaks closer than 125 ms as single bursts. Because of discrepancy between researchers, a variable grouping duration of 15 ms was implemented. It was discovered that a maximum distance of 10-15 ms was consistent in all trials for identifying groups. This duration means that if the time between peaks was no smaller than this duration, activity was present, beginning with the first peak. The user then confirmed that identified points were salient.

Mathematical operations were accomplished with the help of peakdetect.m (Copyright (c) 2001 Tom McMurray, mcmurray@teamcmi.com), and nearest.m (Copyright (c) 2000- by Heekwan Lee (heekwan.lee@reading.ac.uk) Revision: 1.1 2000/04/16), from the Matlab Exchange database.

3.6 Analysis of Results

Latency was the dependent variable, and the outcome of interest. Controllable input parameters were **D** and **c**. The hypotheses were that

a) The algorithm is robust, and there is no dependence between latency and input variables **D** and **c**.

b) Average latency between sEMG onset and motion onset would fall within range of expected latencies for the less affected arm of stroke patients, confirming clinical applicability.

Because limited time confined this study to the design of a wavelet detection method, only two subjects' data were made available for reliability testing. Currently, assessing the independence of measured latency from input variables, and the physiological relevance of those latencies, were necessary steps in determining future applicability of this work. The following analysis is meant as an assessment of the plausibility of the method.

There were two subjects (subjects 1 and 2), with data collected pre and post training (time 1 and 2). Input to the detection method was **D** and **c**, both integers. Input to the supervised classification routine was duration, in milliseconds. Output was first SNR, dependent on the ratio between the noise defined by **D**, and signal content in levels 3:6 (13.5 – 600 Hz, or approximate mean frequencies of 27.4 – 219 Hz with Equation 2.21). The integer identifying the value of the thresholding coefficient, γ, was **c**, which ranged from 1 to 17, corresponding to a γ ranging from 0.8 to 2.4. Each increasing integer value of **c** added 0.1 to γ. The maximum coefficient (**M**) was the maximum noise coefficient found in level **D**. The threshold *th* was calculated by multiplying M and γ, and is therefore linearly dependent and directly correlated to both **M** and γ.

The muscle used for analysis was the biceps brachii, a main contributor to the reach task, functioning as an elbow flexor, and a shoulder flexor [60]. Buneo [15] and Rash [99] both recommend normalizing to speed. Speed of movement was determined by endpoint linear velocity. The recorded task was done at varying speeds, and the effect of speed on the determined latency was considered.

CHAPTER 4

RESULTS

4.1 Outcome

In the reach task, the electromechanical delay prior to motion approximated 147.1 ms (SD 115.1), shown in the highlighted column of Table 4.1. The range of latency time approximates some of the values suggested in the literature, of 100-120 ms (Sub-section 1.2.5). Figure 4.1 is a sample of detection of subject 2 pre MP, trial 3, where input variable s**D**=1, and **c** = 1. All input and output variables for each signal are presented in Table 4.1.

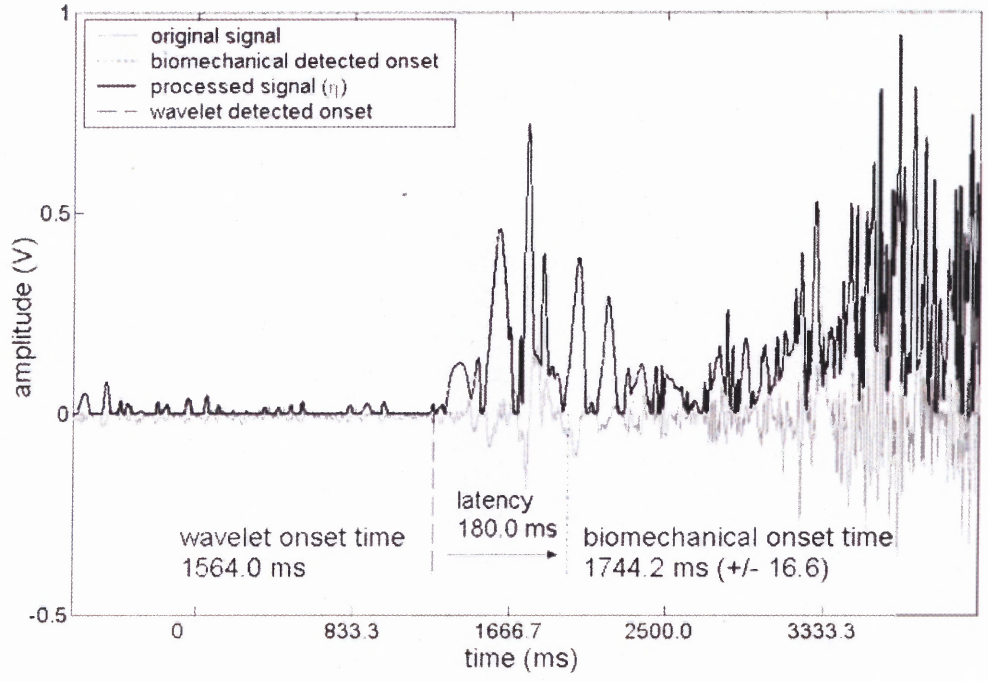

Figure 4.1 Sample onset detection for subject two post trial three, with noise level **D** at 1, and threshold iteration **c** = 1 ($\gamma = 0.8$). The original signal s is in gray.

Examining the raw signal in Figure 4.1, it is difficult to exactly define onset of activity. Greater scrutiny reveals a change in frequency content of the signal. WT can detect this change.

Table 4.1 All Input and Output Variables

	input variables					output variables				
	time (pre or post)	speed (mm/s)	Noise level	threshold iteration	group duration (ms)	SNR (dB)	Max Noise	gamma	threshold	latency (ms)
subj	time	Speed	**D**	**C**	dur	SNR	**M**	γ	*th*	latency
1	1	445.8	2	1	15	13.3	1.26	0.8	1.00	61.3
		529.0	2	1	15	13.1	1.47	0.8	1.17	98.3
		308.9	2	1	15	12.3	1.28	0.8	1.02	117.1
	2	195.8	2	3	15	20.9	0.56	1.0	0.56	176.7
		250.3	2	1	15	13.4	0.58	1.0	0.58	73.3
		250.7	2	1	15	12.0	0.70	0.8	0.56	15.8
2	1	972.5	1	1	15	19.6	0.04	0.8	0.03	121.3
		1103.5	1	1	15	21.7	0.06	1.0	0.06	135.4
		685.8	1	1	15	21.2	0.04	1.0	0.03	180.0
	2	499.2	1	10	15	22.5	0.05	1.7	0.09	476.7
		567.7	1	12	15	22.2	0.04	1.9	0.08	138.8
		552.9	1	13	10	22.5	0.05	2.0	0.11	178.3

4.2 Dependence on Input Variables c and D

The two variables that affected the peaks defining changes in frequency were **D** and **c**, the level defining noise, and iteration for identifying the threshold coefficient, respectively. **D** was either 1 or 2, representing approximate bandwidths of 600-1200, or 300-600.

One indicator of robustness was in how many iterations of **c** the same onset point was found. On average, the same point was found in 5.1 (SD 3.2) iterations of increased/decreased thresholds.

Figure 4.2 shows that latency did not depend significantly on either **D** or **c**. Similarly, Figure 4.3 shows that threshold, of which both **M** and γ are factors, had no effect on latency. Both figures indicate a robust method.

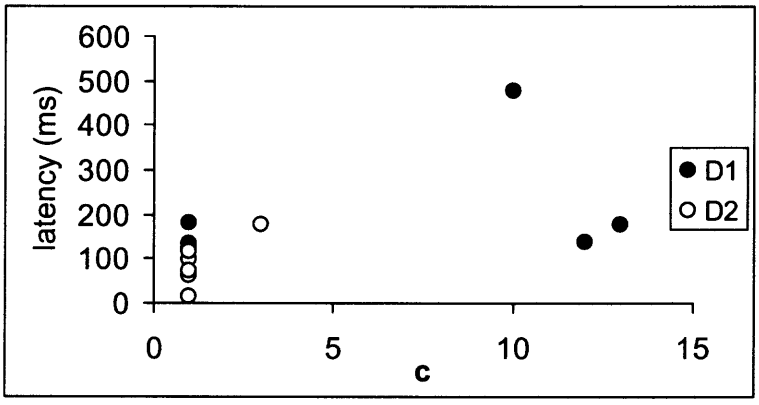

Figure 4.2 Latency with respect to input parameters **c** and **D**. There is no apparent clustering of latency due to either.

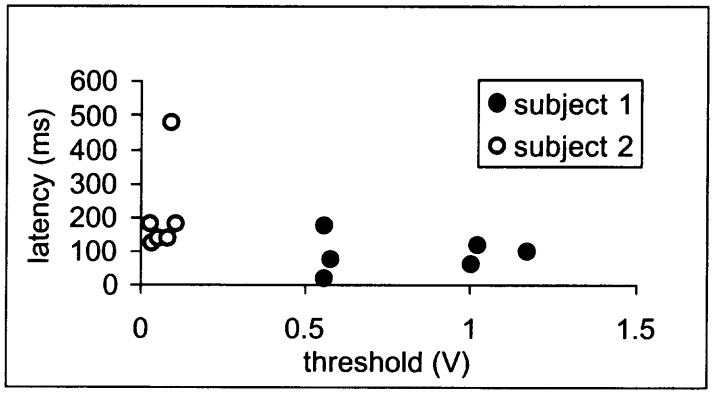

Figure 4.3 Latency with respect to threshold *th*. The choice of **c** influenced the value of *th*, but did not influence latency.

Even though the input variables of **D** and **c** were different for each of the two patients (Table 4.1), the resulting measured latency was independent of input parameters (Figure 4.2) and subject (Figure 4.4).

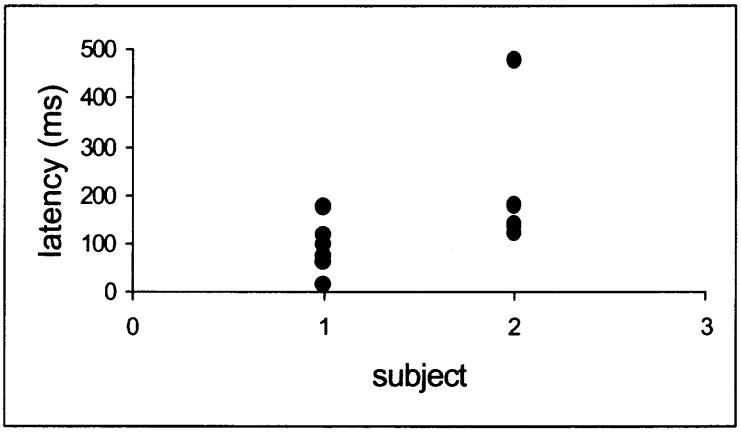

Figure 4.4 Latency with respect to subject. Latency is not significantly related to the subject.

Values of **M** and γ were all dependent on the choice of **D**, and *th* was dependent on **M** and γ. Thus, the variable γ was correlated to **D**, because a **D** of 1 might require a much higher γ than would a **D** of 2 (Table 4.1). Because **D** was related to the SNR of the signal, by inference, the choice of **c** was as well. Thus, it is important to understand that the source of **D** is not arbitrary, but based on the signal it is characterizing. This is expanded further in Section 4.3.

4.3 SNR and D

The SNR is highly related to the noise level chosen. This makes sense, because the purpose of choosing **D** was to establish a bandwidth of noise. The choice of **D** reflects inherent characteristics of signal s. The same **D** could not be used on all data. **D**2 might remove too much information by describing recorded activity as noise, with an **M** that created too high a *th*, making **D**1 the necessary alternative. Choosing **D**1 would measure noise with a lower amplitude, increasing the SNR. Similarly, **D**1 could not be used on all

signals, because the hard thresholding in the next step may then be insufficiently low, due to a small **M** found in **D**1, and **D**2 would then be required.

Figure 4.5 shows that subject one was fit best by **D**2, and subject two, with **D**1.

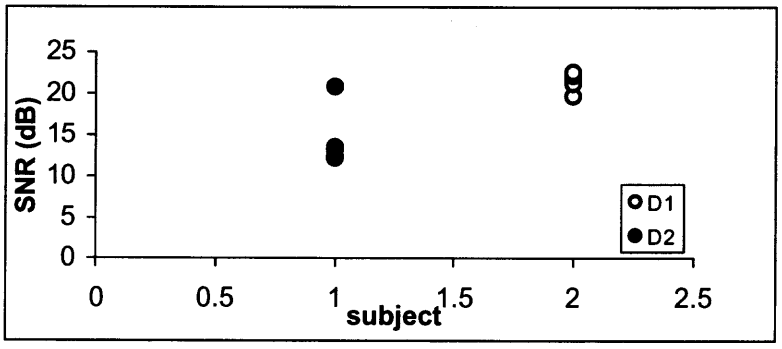

Figure 4.5 SNR used for the analysis for each subject. A noticeable clustering related SNR to each subject, coinciding with a fit of **D**1 to subject two, and **D**2 to subject one.

The standard ratio between noise and salient signal content, the signal-to-noise-ratio (SNR) in sEMG is about 25-35 dB [24]. All trials that used **D**1 returned an average SNR of 21.6 dB (1.1), and all that used **D**2, result in an SNR of 14.2 dB (3.3) (Table 4.2). All SNR dependant variables are thus **D** dependent, visible in Figure 4.5.

It follows that **M** is also very dependent on **D**. **D**1 trials had an average maximum noise coefficient of 0.05 (0.01) V, and **D**2 trials had an **M** of 0.97 V (0.4) (Table 4.2).

To test the dependence of the choice of **D** on signal characteristics rather than the user, the SNR of all signals using **D**1 and **D**2 as the noise band was calculated (Figure 4.6a,b, and Table 4.2). Subject 1 had a slightly higher SNR with either **D**1 or **D**2, and **D**2 created a *th* within usable range for subject 1. There was not enough noise in **D**1 to produce a high enough **M** for efficient thresholding. The reverse held true for subject 2. Figure 4.7 demonstrates how the end result of subject dependence on SNR and thus **D**, did not contribute to latency.

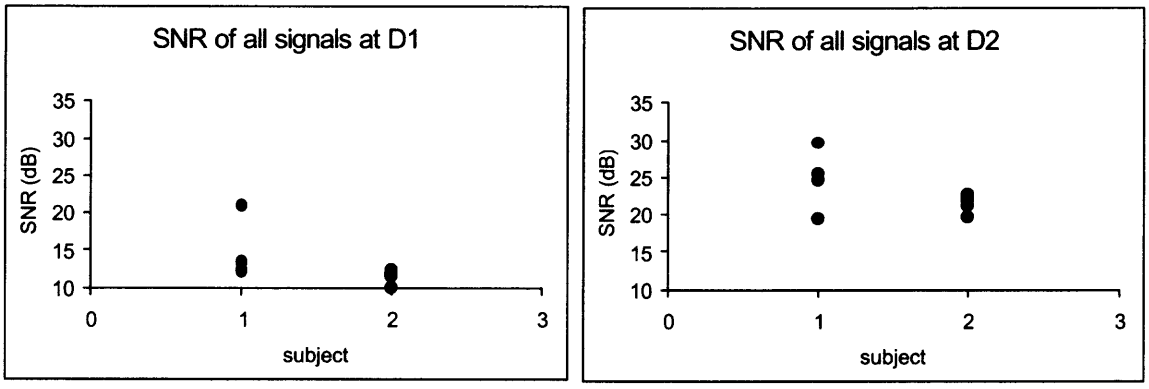

Figure 4.6 a,b: SNR was calculated for all subjects, using both **D**1, and **D**2.

Table 4.2 Input and Output Variables Grouped by Subject and **D**

		input variables			output variables					subject variables		
		Noise level	threshold iteration	grouping duration (ms)	SNR (dB)	Max Noise coeff.	threshold coeff.	Thres-hold	latency (ms)	hand speed (mm/s)	SNR at **D**1 (dB)	SNR at **D**2 (dB)
group		**D**	**c**	dur	SNR	**M**	γ	*th*	ms	speed		
all	$\bar{x}$	1.5	3.8	14.6	17.9	0.51	1.13	0.44	147.74	530.18	12.9	23.3
	σ	0.5	4.8	1.4	4.6	0.55	0.46	0.43	115.1	282.16	2.68	2.89
by subj.	$\bar{x}$	1	6.3	14.2	21.6	0.05	1.40	0.07	205.07	730.27	14.2	24.9
	σ		5.9	2	1.1	0.01	0.53	0.03	135.19	249.5	3.34	3.25
	$\bar{x}$	2	1.3	15	14.2	0.97	0.87	0.81	90.42	330.09	11.6	21.6
	σ		0.8	0	3.3	0.41	0.10	0.28	54.62	129.68	0.76	1.1
by **D**	$\bar{x}$	1	6.3	14.2	21.6	0.048	1.40	0.07	205.1	730.3		
	σ		5.9	2	1.1	0.008	0.53	0.03	135.2	249.5		
	$\bar{x}$	2	1.3	15	14.2	0.971	0.87	0.81	90.4	330.1		
	σ		0.8	0	3.3	0.406	0.10	0.28	54.6	129.7		

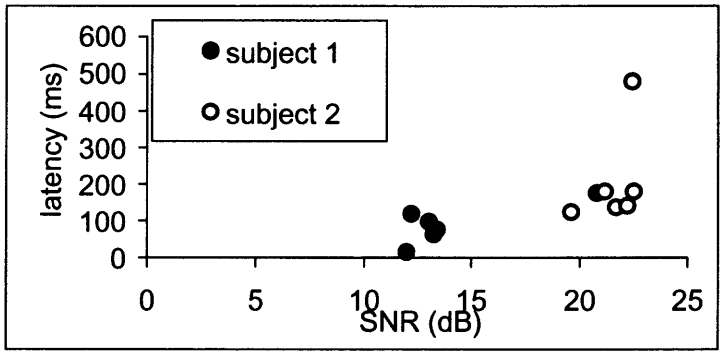

Figure 4.7 Latency with respect to SNR. As relevant as SNR was to the subject and the choice of **D**, the onset latency was not related to SNR, thus by inference, the choice of **D**.

4.4 Dependence of Latency on Hand Speed

In addition to SNR being dependent on the subject, each subject had a chosen linear endpoint speed, calculated from a marker on the second metacarpal joint.

Subject 1 had a mean endpoint speed of 330.1 mm/s (SD 129.7), and subject 2, 730.3 mm/s (SD 250.5). There is a strong relationship between subject and speed, and subject and latency time. However, there is no clear association between speed and latency (Figure 4.8). Commonly accepted relationships between speed and latency would imply that there would be increased latency with decreased speed, but that was not evident in the data.

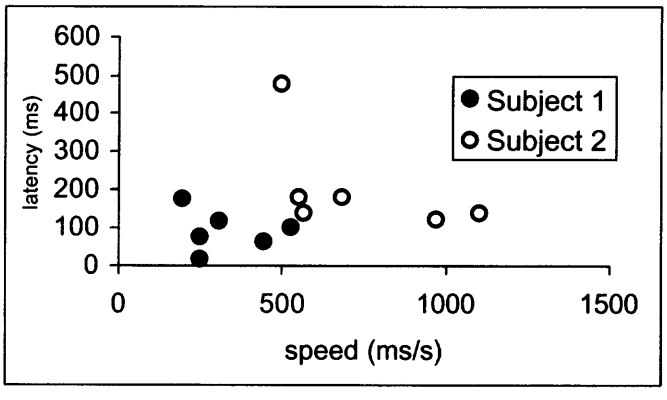

Figure 4.8 Latency with respect to speed. Each subject had a different mean speed. Neither speed nor subject had influence on latency.

4.5 Dependence of Latency on Subject and Time

Although the input variables were subject dependent, predictable by SNR and **D** being unique to each subject, there was no clear relationship between subject and latency. (Figure 4.9, Table 4.1). Time pre and post also demonstrated no effect on latency. However, many factors are closely associated with subject, such as SNR and thus **D**, followed by **c**. Many more subjects would need to be collected to verify the dependence of latency on subject, as opposed to other input variables, because the input variables so strongly related to the subject. However, since neither **D** nor **c** were clearly indicative of the latency, nor was the SNR responsible for differences, it is reasonable to see that it was the subject that was relevant, as would be expected physiologically.

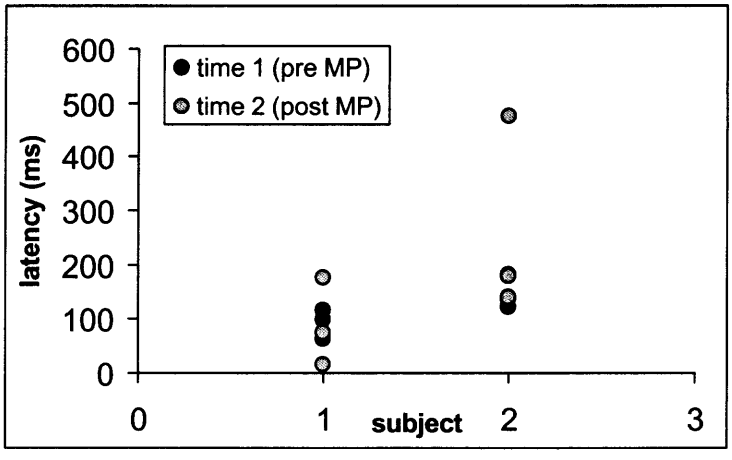

Figure 4.9 Latency with respect to subject. Although latency was not effected by pre or post MP, each subject had slightly different mean latencies, with subject two having higher latencies than subject one.

CHAPTER 5

DISCUSSION

5.1 Outcome

Wavelet analysis was successful three ways.

a. Latencies between onset times detected in sEMG and motion initiation times were physiologically validated. An average latency of 147.7 ms (SD 115.1 ms) corroborated with normal latency values in the less-affected arm, of 90 - 120 ms [34; 59] and did not exceed 163 ms [21].

b. The method is robust. Output latencies were not dependent on input variables, nor were input variables unique.

c. The method identified initiation of activity, not where activity crosses thresholds, but where changes in frequency contributed to motion.

5.2 Latency Error

Flanders 1992 [44] found that there are 100 ms on average (up to 120 ms for slow motion) between agonist burst and movement onset. Hudgins [38] indicated that in the 100 ms prior to onset, a great deal of information is present. This was evidenced when the average onset for the less-affected biceps brachii was found to be 147.7 ms (SD 115.1 ms). This latency value is somewhat larger than suggested in the literature, with an even greater standard deviation. This difference may be due to the ability of the wavelet to detect changes more deeply into noisy regions prior to the start of motion. Furthermore, the variability in the literature means that there is no set value of what the latency should be, with differences up to 100 ms across sites for the same measurement. Therefore, an approximation had to be used. To circumvent this, future research on this method will

include more precisely measured onsets of motion. Data will be collected on normal subjects, for whom latencies of 100 ms is accepted as standard

Errors in measuring latency more probably lie in the video data of motion, in that the motion data was captured at 60 or 120 Hz. If the start of movement was misidentified one sample before or after the actual start, an automatic error of 16.6 ms or 8.3 ms would occur. In contrast, the sEMG data was sampled at 2400 Hz, and a single sample error would contribute only 0.4 ms of error. With the limited time available, not enough data was tested in order to exclude video capture quality as a source of error.

In the current study, the electrode placement was less than optimal. DeLuca 1997 [32] indicated that if electrodes were placed poorly, meaning more than 2 cm apart or too close to the innervation zone, the result of the differential amplification between the electrodes may extend the frequency content of the sEMG signal into higher ranges (Figure 1.7). The maximum frequency in proper testing conditions should not have exceeded 200 Hz. However, because the second electrode was placed on a bony landmark, the power spectrum of the EMG looked like Figure 1.8. The frequency bands that would ordinarily be only noise, i.e., bands 1 and 2 with pseudo-frequencies greater than 250 Hz, still contained information related to the motion. For this reason, **D**1 was a frequently required level for noise identification. With that in mind, the correlation between trials requiring **D**1 assignment as noise may relate more to electrode placement on that day, than to the subject. This relationship will be explored in the future, on studies with properly applied electrodes.

Canning's [19] findings show that slow motion due to loss of dexterity was often accompanied with much faster muscle activation, possibly because the subject was racing

to overcompensate poor control. Figure 4.8 shows that slower speeds may have slightly lower latencies that could occur with faster recruitment of MUs. Evidence of shorter latencies accompanying slower speeds is in contradiction with commonly accepted principles regarding motion, that indicate that slower speeds should have increased latencies. In the cases studied here, the subjects chose their own speeds. It may be that each subjects' speed was similar with respect to each patients' maximum ability, in effect, normalizing their speed. In many studies, patients are requested to execute the task as fast as possible, stressing a system to its maximum ability, thus eliminating some variability, each patient being forced to perform at the maximum speed attainable. Currently, testing of the method is being done on subjects pointing at what they consider a 'fast' speed.

5.3 Input Parameters

Isolating one level for calculating SNR is viable, even though it is a user dependent criterion, because the SNR value is a self-reflective value used for weighting a synthetic signal. In addition, white noise being sufficiently represented by one level, because it is uniform throughout a signal [41; 115].

The method was successful in that modifiable input parameters **D** and **c** had no bearing on the measured latency. SNR appears to be completely dependent on the user's choice of **D**, except that the choice of **D** is dependent on inherent signal quality, not user discretion. The gray series of Figure 4.5 indicates which of the signals used **D1** as noise when calculating SNR. If the SNR from **D1** was too high, the implication was that **D1** did not have sufficient content for useful calculations, and a high threshold would be

required to clean up η'. If this was the case, *th* calculated with all possible values of γ is still insufficient to remove noise, so **D2** would have been chosen instead. Similarly, if the SNR was so low when calculated with **D2** because information resembling activity was found in **D2**, **M** would be high, and the resulting high *th* would remove too much information. In all signals tested, there was no room for subjectivity in choosing **D**.

Still, activity was only present in **D1**, a high frequency band, due to poor electrode placement, which stretched the power spectrum above the standard 250 Hz limit for sEMG in reach data, as in Figures 1.7 and 1.8.

Regardless of SNR, an indication of robustness was that there were a number of iterations of thresholding that contained the same onset point, even as high as 10 iterations. This meant that the user could chose a large range of **c**, and still obtain the same onset point. More test signals would enable the proper and statistical exclusion of **c** and **D** as dependent variables.

5.4 Detected Onset

The same values of γ removed less information when there was a high SNR than lower SNR. If SNR is high, the value of γ was not that critical, and can be large [77]. A low value of γ is preferable with low SNR, because a high *th* might remove useful data. Similarly, higher SNRs required lower thresholds to remove noise, because the definition of activity over noise is clearer. Future studies will help indicate how strongly latency correlated to each subject, due to that subject's SNR and speed.

The points identified as onset are not always obvious to the user, and may appear earlier or later than increases in amplitude. It must be stressed that true onset is evidenced

in changes in the frequency domain, as well as amplitude changes [39]. In fact, within the transient burst of myoelectric activity accompanying the onset of sudden muscular effort, there is significant temporal structure to encode information important for pattern discrimination (Hudgins in [39]). Standard methods (described in Section 5.6) might have simply identified the increase in amplitude, but those occasions may be after the motion initiation has been marked. Here, not only changes in amplitude, but also changes in frequency content, are identifiable. (See Figure 4.1).

With WT methods, it is the characterizing array of high coefficients at each point in time that is being modified by input parameters, not an altered version of the original signal. Manipulating a characterizing array maintains integrity, because events in the array can be analyzed independently from the signal, left in its original form. For instance, the value of the interval to define activity bursts between coefficient peaks is adjusted, instead of smoothing the signal until topological patterns emerge.

5.5 Comparison to Other Wavelet Methods

The amplitude of surface EMG is often used to measure muscular effort [23]. In any description of what constitutes muscle 'activity', there must first be a method to find where in time the amplitude of the signal is sufficient to be considered 'active', and 'inactive', for any determinations regarding the length of that activity. Currently, there is no specific rule for choosing processing techniques. Consequently, the relationship between EMG and a parameter under analysis (such as tension), is not unique across experimental sites even with carefully controlled environments [89].

Wavelet transform methods may provide an alternative, because it is a method applicable to a number data types, both cyclical and single motion, and it does not only consider amplitude. Detection with the routine developed here is not site-specific. This same wavelet method, with a different wavelet, was based on Merlo's [77] data that was recorded in a completely different collection environment. The input parameters were still similar in both methods.

Several methods of using wavelets for feature extraction with CWT do exist [4; 26; 38; 43; 55; 63; 81; 88; 126]. Flanders [43] used peaks in coefficients from CWT with db4 to identify maximum portion of sEMG bursts, but not to find their onset. Khalil's method [63], while established for detection, was tested on uterine EMG, a highly distinct signal, and only used wavelets for classification, for which the Coiflet wavelet was found to be valuable. Hsu [55] used coefficient maxima from a wavelet that was the first derivative of a Gaussian function, but he tested it on hyperspectral data. Arikidis [4] found local maxima with CWT, and thresholded coefficients at various scales. However, this was used on MUAPs from indwelling wire electrodes. For classification, Arikidis found salient events to occur when maxima across scales marked the same temporal event. In the method developed here, a single array of maxima was tested, which found the highest maxima in all scales at each point in time, a simplified approach to Arikidis' method. Future work with this method may include using Arikidis' method for assessing importance of detected events.

Englehart [38] successfully extracted events from sEMG using a Coiflet 4 wavelet. However, the sEMG signals were from four electrode channels, and, he found that steadily changing data was better assessed than transient data, such as studied here.

Panagiotacopulos [88; 126] does not divulge his method, however, he indicated that the db8 and coiflet3 were among the better wavelets for activation detection, with precise and repeatable detection. Mochimaru [81] used a Daubechies 20 wavelet for denoising and a Coiflet 24 wavelet for isolating two concurrent EKG signals. Thresholds for various frequency bands were based on the standard deviations of the coefficients at each level. Maxima were then treated with classification routines, such as PCA and Lipschitz exponents, which relate to the direction of coefficient peaks or troughs. While using the same basic concepts as applied here, Mochimaru moves into more complex statistical analysis, because his signals contain two salient traces. Detection of onset was not of interest, just the differentiation of waveforms.

Conforto [26] used a statistical detector on a wavelet filtered signal. However, in addition to removing levels containing only signal noise, Conforto uses self-referential filtering of coefficients in wavelet subbands to effectively remove motion artifact. With more time, an advanced subband filtering such as developed by Conforto could be applied to the data tested here, instead of simply weighting to zero bands that contain recognizable noise.

Zennaro's [126] decomposition was developed for detection of slight movements in multichannel fine-wire long term simulated EMG. There is prior knowledge of which segments contain activity. After filtering and segmentation, wavelets are used twice, once for clustering the segments, and again for classifying those segments. Templates and best-matches to those templates are maintained throughout. Again, while Zennaro uses wavelets for classification, the detection of active EMG over non-motion, is not his goal.

Currently, only Merlo's method, tested on synthetically induced signals in a non-living limb [77] was designed to accomplish similar needs as at KMRREC, where this study was performed. Wavelet detection methods have a high quality of discrimination, and the accuracy of other WT methods is not going to be challenged.

The method designed here could be upgraded with the filtering, parameter extraction techniques, and classification routines from all of the above methods. Still, the method used here was far simpler, easier to construct, and able to identify the same point of onset just as well as the more complex methods. If the method were upgraded, the benefit would be in the automation of analysis, and is under development.

While the SNR values that Merlo found and used for selecting γ ranged from 8 dB to 12 dB, and the SNR values in this method were higher, ranging from 10-27 (Table 4.1), they were still similar in value. In fact, the ranges of γ paralleled the different range of SNR.

5.6 Standard Methods

The onset of sEMG activity is often estimated from biomechanical data using one of the following techniques [35; 42; 54; 88; 106]:

a) A combination of amplitude and duration thresholds. Amplitude thresholds are usually based in some way on the noise level in a pre-movement or maximum voluntary contraction signal [54], and is often combined with a durational threshold [91; 121] that defines how long activity above threshold must last to consider bursts as activity.

b) Spike characteristics, such as the rate of rise [10] [39], turn counts [89], the number of turning points in the signal that are separated from surrounding turn points by amplitude changes [23] or an amplitude difference greater than a prescribed number of volts [42], are often used to define onset with less smoothing than techniques in a. Other variables include turns/second, amplitude/turn, ratio of those two, as well as measures of minimum spike

duration [37]. These methods require significant filtering to reduce noise. Spike counting is similar to turn counting, with the additional factor of the density of spikes per unit time [124]. These parameters are time consuming to define and very subjective to the test site.

c) Combinations of these methods include statistical predictions of the density of peaks required to consider as activity, or multiple threshold measures derived from densities, amplitudes, or rate of rise coupled with amplitude [96]. Signal content methods may include SNR based thresholds, or whitening filters [104] often combined with statistical detection methods [107]. Topologically related thresholds are often based on Gaussian parameters, such as multiples of standard deviation from a raw signal's Gaussian mean or silent region [1; 54; 78], or ratios of standard deviations of active to that of silent periods.

d) More advanced pattern matching [76] can be done with cross correlations, neural networks [22; 49; 84], or fuzzy logic.

Absolute amplitude of EMG from day to day is quite variable. Normalization techniques have included setting a maximum voluntary contraction (MVC) to 100% amplitude, or using baseline or pre-movement signal content as a reference. These second signals are just as variable as the signals to which they are being applied. Due to the rate of muscle activation or the presence of artifacts, there is so much variability across EMG signals, including MVCs or resting signals, simple thresholding methods are inappropriate [3; 96]. Thresholds do not necessarily exclude transients and noise, which may be of similar amplitude as salient data, but not the same frequency content [48]. If the signal changes from inactive to active in the same location that noise is present, with no amplitude increase and only a subtle frequency change, the event will appear to be under threshold, and onset will not be noted unless further constraints are added [116]. Still, because of their mathematical simplicity and availability, thresholding techniques are in common use.

Thresholds are chosen based on within-signal details such as maximum amplitude [36] for EKG, [117; 121], amplitude distribution and peak statistics [3; 12; 32; 78], values relative to MVC or baseline noise [45; 91], or combinations of descriptive parameters such as amplitude and spike durations [37].

The definition of what is baseline, and thresholds for what constitutes deviation, is where methods differ. A sample rise above baseline of a smoothed signal is shown for subject two, trial three, pre training (Figure 5.1). A great deal of information is lost at the expense of preserving the shape of the signal [96]. Compare this with the coefficient array characterizing the signal in Figure 4.1.

The wavelet method is a substantial improvement over traditional methods. Particularly in signals with higher SNRs, the amplification of the signal content over noise content occurring when the maximum coefficients are taken, improves the precision available. Patterns become pronounced. This clarity is not obtainable in standard methods

One way to reduce the variability present in standard methods is to eliminate secondary reference signals, such as noise signals or maximum voluntary contractions (MVC). Using MVCs for normalizing the signal or to extract threshold parameters is still common, despite research to prove that MVCs are highly unreliable [13; 39; 67; 79; 99].

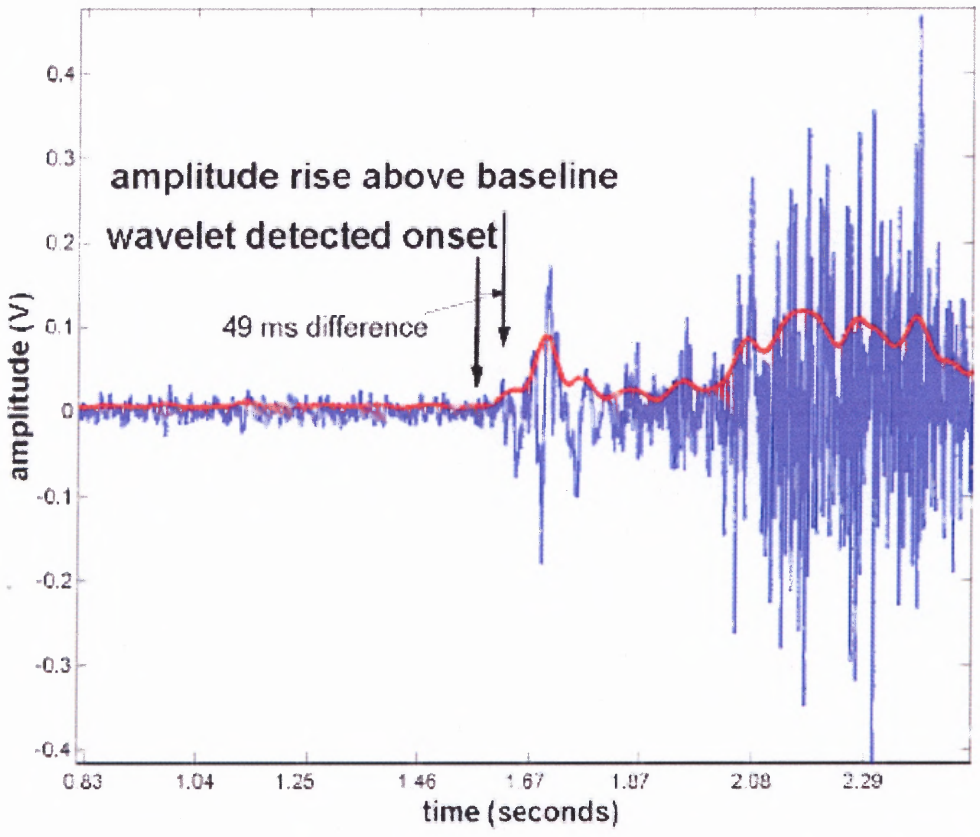

Figure 5.1 Sample of standard smoothing techniques on subject two, trial three, post training. The signal is rectified then smoothed with 15 Hz low pass filter. The point of deviation of the smoothed signal from baseline is often considered onset. Note the difference between the point in time in which the amplitude of the smoothed signal becomes greater than baseline and the precision of onset detected with WT methods in Figure 4.1.

An MVC is collected during isometric activity, and a noise signal would be collected while the subject is still, or identified as inactive regions in a test signal. Both signals may not contain information that may be present during dynamic motion. The WT method is completely self-referential, requiring no prior knowledge or second signal for parameter definition. The method here reduced errors that would have been introduced by different gains and testing conditions, by using a noise signal extracted from the collected signal, instead of an independently collected or identified noise signal.

Noise for the SNR calculation was defined as one level from the decomposition with DWT that showed uniform activity throughout. While the wavelet method considered properties such as SNR inherent in the signal for defining threshold parameters, the thresholds were based on what was present in each signal, the parameters were not unique to each signal or user, and were robust.

Furthermore, the results illustrated that noise bands defined with DWT for SNR calculations is a plausible alternative to using signal collected separately. A subject need not spend additional time while instrumented, in order to collect a second calibration signal.

When different sites attempt to apply other algorithms, parameters must be specialized to every site's data sets. Methods are not universally applicable. Because so many thresholds have been proposed and none of them consistently accurate, variable threshold detectors have also been applied, with similar error to automated ones, and with a great deal of supervision required [121]. Error components across tests in these methods are also rather high, requiring supervision and adjustments [70], and further statistical methods [3; 37; 70], with limited success. A number of findings indicate that standard threshold methods are about as repeatable as simply picking off activation points by eye [21; 35; 42; 57; 57; 96; 106] with at most 85.4% success [70], or with coefficients of variation as high as 20% [121]. More complex statistical methods based on turn counts and the like are just as poor, with reported incidence of the same millisecond being chosen occurring in only 30-36% of chosen instances across all examiners, and 47-56% within examiners [54]). Even commercial algorithms using more advanced methods of automatic detection, such as statistical detectors, neural networks or deconvolution, have

produced similar or greater variabilities. Bogey et al found detected onset time differing within one algorithm by 5% (3.1%) [11]. However, each percentile is in quantized units of 100% gait cycle, which can reach 160 ms per unit. Others have measured erroneous activation placement occurring between 5% [25]- 20% [36; 37; 52; 53; 54; 64] of the time. Variability across methods can be up to 100 ms [88].

Additionally, the standard methods are not usually related to the neurophysiology of EMG, but rather the apparent shape of IPs recorded with sEMG. The thresholds that have been used are somewhat arbitrary [121] and weighed down with time-consuming and subjective adjustments [95; 96; 120]. While thresholds were used here in the wavelet method designed in this study, the precision of the parameters defining threshold were not critical, as evident in Figure 3.1. The robustness of the WT method means that it can identify the same sample (an error of only 0.8 ms) with a number of thresholds, sometimes up to 10 iterations. Furthermore, the exact point of onset was temporally relevant to the raw signal, with no lag or change in time unit due to quantizing or smoothing.

More advanced methods including neural networks, statistical counting or fuzzy logic all require a significant amount of preprocessing and filtering, and are complicated to apply [49], requiring layers of constraints and inputs. The core of the WT method used in this paper encompasses just a few lines of code that can easily be shared with Matlab users (bolded lines in Appendix A.1). Neural Networks [22; 49; 84] and pattern matching [71; 76] for biosignals require advance knowledge of the form of training sets of previously identified activity. This is similar to defining an appropriate wavelet, but is more complex mathematically, requires a computational training period for teaching a

successful pattern, and is somewhat heuristic. No prior knowledge of the pattern of the signal is required in the wavelet transform detection method. For instance, the Daubechies 6^{th} order wavelet has been applied to multiple types of EMG – both surface and fine wire, and on differently sized muscles. Furthermore, more advanced shape matching filters still may have a low incident of success, as low as 88.7% [105], not much better than simplistic thresholding.

Another difficulty with these other methods is that prior to applying the analyses, the signals must go through a series of signal processing routines [99]. It is important to be aware that almost all of these methods apply some combination of rectification [23], RMS calculations [16], enveloping, filtering [54], smoothing with moving window integration [3; 14; 88], signal processing filters and smoothing [104] or normalization techniques, that introduce potential variability [61; 72]. A window of integration is often as large as 100-200 ms [99]. The signals then being searched for activation have significantly reduced time resolution. In contrast, a signal assessed with a WT does not need to go through any processing before locations of frequency change are identified. It is the array of characterizing coefficients that undergoes manipulations of thresholding and burst detection, not the original signal.

Furthermore, as explained in Subsection 2.4.1, filtering techniques are based on a series of sine and cosine waves, of varying frequencies. A single MUAP does not match the model of a sine wave, but a wavelet can be designed that does. Daubechies is one wavelet family similar to a MUAP [56].

Still, because of the ease of development in clinical laboratories, threshold techniques following standard filtering and smoothing methods are still in use, despite

their inefficiency and the decrease of performance of traditional methods during highly dynamic motion [106]. The advanced methods designed to combat insufficiency of thresholding routines are also not really used in a clinical environment due to their complexity of design and application. Recently however, there was a shift to wavelet methods. In the course of this research, it was observed that a number of researchers formerly involved in the development of some of the advanced techniques described below, have adopted WTs as an improved methodology. Conforto [25] had developed a dual threshold in 1999, and it was upon her dissertation that Wilen [121] based a previous method of detection. Khalil [64] was using probability density functions as late as 2000, and, like Conforto, is now using Wavelets for detection.

5.7 Supervised Classification Subroutine

The WT method successfully identified bursts. It is up to the user's discretion, based on knowledge of the motion at hand, to define how many bursts incur activity, based on the amplitude or proximity of spikes. The last step of the WT method involved the user confirming the automatically detected regions of activity. The user does not actually choose or confirm the onset points, just which bursts can be grouped together. While this is not optimal, several researchers have concluded that some level of heurism is required for most EMG non-statistical activation detection protocols [74; 106; 110; 117]. Even "training" datasets for neural network methods are user defined. These datasets are often based on initial samples of activation selected to represent a pattern, which are obtained with a similar heuristic method as used here, except often with only one amplitude threshold. Even in advanced recent developments, supervised classification is still a

classification is still a standard and acceptable final stage after the actual detection [40; 126], because the gaps between onset will vary with movement type and speed. However, a heuristic classification method is not using wavelets to their full potential, which have been used for such purposes [5]. Due to limited time, the development of a statistically sound classification routine with wavelets was not yet accomplished.

CHAPTER 6

CONCLUSION

A wavelet method for determining onset in sEMG of the biceps brachii in pathological data was successfully designed. The method is superior to methods in common use, in several ways. It identifies sEMG onset in the same temporal resolution as the original signal, because there is no lag ordinarily introduced by Fourier based filtering or integrative smoothing. This method is completely self-dependent, and does not require either a second noise signal or MVC signal for calibration, reducing both the time spent with a patient, and data storage needs. A significant advantage over complex methods that are under development elsewhere, is that it is simple to design and apply to many different EMG collection environments. The maximum coefficient array is a very manageable characterizing array that can be input into more advanced grouping techniques.

Furthermore, the method is robust, in that user inputs are not very critical to identification. The inputs are not arbitrary like many thresholding algorithms, and depend heavily on a signal's own characteristics, such as SNR.

Future work would include a validation of the method by testing the method on a larger sampling of subjects collected with properly placed electrodes. Periodic data with undefined periods of inactivity, such as gait data, where secondary 'inactive' signals are unavailable, is currently being assessed. Other wavelets, such as the Coiflett, which have similar properties required for this method, will also be tested. Advanced classification

routines using coefficient maxima more advantageously will also be explored, including the automation of the selection of **D** and **c**.

Publications on this method have been submitted to the 2004 Gait and Clinical Movement Analysis society meeting, entitled "Using Wavelets for the precise detection of Muscle Activation in EMG of the Biceps Brachii"

APPENDIX A

MATLAB SOURCE CODES FOR DETECTION WITH WAVELETS

This appendix gives the source code for steps 1-8 of the methods described in Section 3.5, entitled WaveletDetect.m. Input parameters are the file, noise level **D**, and threshold coefficient iteration **c**. Output is the maximum coefficient array η_c, on which subroutine the supervised classification routine, NuonDetect.m, is performed. The subroutine for defining the pFa associated with each scale, ranging from $1:2^m$ entitled pseudofreq.m, is attached. Bolded lines indicate the core operations, and the rest is data handling.

A.1 WaveletDetect.m

```
% WaveletDetect.m
% This program uses functions in the Matlab Wavelet Toolbox, and was based on
% Finprog.m, created by AM Petrock 11/20/2001.
% Taking input signal Sa, the program will perform DWT, bandlimit the signal
% from a limited number of reconstructed detail arrays, feed the signal into CWT, build
% an array of the maximum coefficients at each point in time, and threshold that array
% by a threshold defined by the maximum coefficient in a user defined level of noise
% from the DWT and a coefficient gamma (identified by threshold iteration integer c)

%Test for existing input array
if exist('Sa') == 0;
   error('please load an EMG file, and save the appropriate data column as Sa')
end

subjlabel=input('type file name, append side: ','s');

%Set common input variables
disp('set for reach data')
maxscale=7;          % the max m scale is 7 for most data
reconstruct=1;       % an option for a bandlimited reconstruction
wave='db6';          % db6 for reach task data
sf=2400;             % the data was sampled at 2400
DefNoiseLevel=2;     % Default D is not set, because it needs to be modifiable
DefArtifactLevel=3;  % Hi cuttoff for bandlimiting s is set for approximately 200 Hz
```

```
DefNoiseLevelD=6;     % Lo cutoff for bandlimiting s is set for approximately 27 Hz
Defcn=3;              % Default threshold iteration coefficient should be 3 (a gamma of 1.0)

% clear local vars:
clear ('mincD','maxcD','A','D','cD')

% run to calculate relationship between scales a and frequency
pseudofreq

% Justify along zero mean line to remove baseline shift
ToJust=1;
if ToJust==1,
  S= Sa-mean(Sa);
else
   S=Sa;
end

% allows the option of using the original S or a bandlimited, filtered version.
if exist('reconstruct') == 0,
reconstruct=input('to asses full bandwidth of 1:maxscale, 0; limit bandwidth, 1; ') ;
else
   disp('a limited bandwidth will be selected')
end

%%%%%%%%%%%%%%%%DWT %%%%%%%%%%%%
% Perform DWT wavelet decomposition into coefficients C and length scalar L
[C,L]=wavedec(S,maxscale,wave);

MaxS = max(S);

% a counter for plotting reconstructed arrays
OddOrEven=rem(maxscale,2);
if OddOrEven==0,
   num=0;
else num=1;
end

% Recompose 'filtered' forms
for i=1:maxscale
   A(i).a = wrcoef('a',C,L,wave,i);
   D(i).d = wrcoef('d',C,L,wave,i);
   cD(i).cD = detcoef(C,L,i);
end

% Plot all the reconstructed waves based on the approximations and details
figure
```

```
subplot(maxscale+1,2,1), plot(t,S), ylabel ('S'), title([wave, ' wavelet decompostion']),
axis([0 LS/sf min(S) MaxS]);
subplot(maxscale+1,2,2), plot(t,S), axis([0 LS/sf min(S) MaxS]);
xS=xlim;
yS=ylim;
for k=3:2:maxscale*2+1
   subplot(maxscale+1,2,k),
   plot(t, A((k-1)/2).a),
   axis([xS yS])
   xlabel(['Approximations ', num2str((k-1)/2), '; Low pass filter'])
   ylabel([' pF ', num2str(scal2frq(2^((k-1)/2),wave,1/sf),'%3.0f')],'Rotation',0,'Position',[-
1.5 0])
end

for h=4:2:(maxscale*2)+2
   subplot(maxscale+1,2,h),
   plot(t, D(h/2-1).d),
   axis([xS yS])
   xlabel(['Details ', num2str(h/2-1), '; High pass filter']),
   ylabel(['level ',num2str((h-2)/2)],'Rotation',0,'Position',[-1.5 0]);
end

% identify noise level D
NoiseLevel=input(['choose noise levels of D; (default = ',num2str(DefNoiseLevel),';
hit enter):']);
if isempty(NoiseLevel)
   disp(['default level to define noise set to ',num2str(DefNoiseLevel)]);
   NoiseLevel = DefNoiseLevel;
end

% One level of D is used at this stage. Further developement may include all residual
levels
lNL=length(NoiseLevel);
noise=D(NoiseLevel(1)).d;
disp('that level is approximately this frequency band of noise:')
pNoiseLevel = scal2frq((2.^NoiseLevel), wave, p)

if reconstruct==1,
   % Low level artifact (norm < 25 Hz) AND high level noise
   % (norm >250Hz) in "noise"
   disp('limit range of analyzed signal; reconstruct set to 1');
    ArtifactLevel=input(['choose lo pass level of D. This WILL be included (default =
',num2str(DefArtifactLevel),')' ]) ;
   if isempty(ArtifactLevel)
      ArtifactLevel = DefArtifactLevel;
      disp(['Default lo pass cutoff (high end of frequency band) set to level
```

```
',num2str(DefArtifactLevel)])
   end
   NoiseLevelD=input(['choose upper cutoff noise level of D (high pass). This WILL be
included (default = ',num2str(DefNoiseLevelD),')']) ;
   if isempty(NoiseLevelD)
      NoiseLevelD = DefNoiseLevelD;
      disp(['Default hi pass cutoff (low end of frequency band) set to level
',num2str(DefNoiseLevelD)])
   end

   % include only bandpassed signal to cwt
   newSrange=[ArtifactLevel:NoiseLevelD];
   S=D(newSrange(1)).d;

   for ns=2:length(newSrange)
   S=D(newSrange(ns)).d+S;
   end
   disp(['signal to be analyzed summated with D levels ',num2str(ArtifactLevel),' to ',
num2str(NoiseLevelD)])

else % if only a bandlimited signal is to be used, and reconstruct was 0,
   S=(Sa);
   disp('original signal used for SNR calculation and CWT; reconstruct was set to 0')
end

%%%%%%%%%%%CWT %%%%%%%%%%%%%%%%
% Perform and plot the continous wavelet transform
figure,subplot(2,1,1), plot(S)
subplot(2,1,2), ccfs=cwt(S,1:2^maxscale,wave,'plot');

%%%%%%%% detection using CWT and Merlo, with Bendat %%%%%%%%%%%%
% obtain the frequencies that generate the strongest coefficients at each point in time
for t=1:LS, [etaSig(t),Isig(t)]=max(ccfs(:,t)); end
pfISig=scal2frq(2.^Isig,wave,p);

% find the maximum coefficient found in the noise signal
ccfsN=cwt(noise,1:2^maxscale,wave);

% get M (max of nu of noise)
for t=1:length(noise), [etaN(t),IN(t)]=max(ccfsN(:,t)); end
M=max(etaN);

%[SR]=SNR(Sa,noise); The s' and noise is used for SNR, and
% the resulting thresholded NuSig is used for analysis of onset
signal=S;
[SR]=SNR(signal,noise);
```

```
disp(['SNR is ',num2str(SR)]);

% Threshold iterations define gamma (threshold coefficients)
c=0;
for etas=0.8:.1:2.4   % changed from 1.8 to 2.0 on 7.3.03
   c=c+1;
   th(c)=etas*M;
end

gamma=[0.8:.1:2.4];  % 1.6 is the value most preferred by the designer, Merlo 2003;
                     % ranged from 1 -2.4. 1 wasn't low enough with these SNRs

% Anything below threshold, (defined by SNR of the original filtered signal, not max
% ccfs) is 0; (Merlo, 2003)
figure
for g=1:c,
   temp=etaSig-th(g);
   for i=1:length(temp),
      if temp(i)>0,
         etaSigTH(i)=temp(i);
      else etaSigTH(i)=0;
      end
   end

% Each row of the figure is eta_c, the iteratively thresholded eta'.
eta(:,g)=etaSigTH';
subplot(c+1,1,1),plot(Sa),set(gca,'YTickLabel',[]), ylabel('Sa'),
subplot(c+1,1,g+1), plot(eta(:,g)), set(gca,'YTickLabel',[]), ylabel(num2str(g))
   if reconstruct==1;
       subplot(c+1,1,1), title(['eta: SNR from noise of level D',num2str([NoiseLevel(:)]'),
'compared to S built of levels D',num2str(NoiseLevelD),':D',num2str(ArtifactLevel)])
   elseif reconstruct==0;
       subplot(c+1,1,1), title(['eta: SNR from noise of level D',num2str(NoiseLevel(:)), '
compared to completely rebuilt S, from full range of D1:D',num2str(maxscale)]),
   end
end

% FOR INPUT TO NuOnDetect.m
cn=input(['choose an eta_c (default = ',num2str(Defcn),', c = 1.0) ):']);
if isempty(cn)
   cn = Defcn;
   disp(['default cn set to ', num2str(Defcn), 'assuming a gamma = ', num2str(gamma(cn),'
and th of ',num2str(th(cn))])
end

% Input into subroutine NuOnDetect.m
```

A.2 NuOnDetect.m

```
% NuOnDetect.m, for use after WaveletDetect
% This is the secondary classification routine that identifies onset by peaks in the final
% thresholded coefficient array eta_c, that occur after a duration ms of no spikes

disp('use peaks plot to help determine onoff bands')

% set default duration check
Def_durms=15;
durms=input(['ms range for activity detection ',num2str(Def_durms),': ']);
if isempty(durms)
   durms = Def_durms;
   disp(['ms range of activity detection was set to ',num2str(durms),' ms']);
end
dur=5*round(1200*durms/1000);
disp(['current allowable gap:',num2str(durms),' ms,',num2str(dur),' frames.']);

% clear local variables
clear('dhn','fdhn','fdhn_off','startEnd','endOff','on','off','w','z','on_count','off_count')

% differential of hn gives you +1 for onset, -1 for offset of every peak.
for i=1:length(nu), if eta(i,cn)>0, hn(i)=1; else hn(i)=0; end, end

dhn=diff(hn);
fdhn=find(dhn>0);
lf=length( fdhn);
fdhn_off=find(dhn<0);
lf_off=length(fdhn_off);

% starting point hunt:
startOn=find((fdhn)>dur);
startEnd=find( fdhn > length(dhn)-dur );
if isempty(startEnd),
   [val_startEnd,startEnd]=max(fdhn);
end
startEnd(1);

% endpoint hunt:
durms_off=durms *1.5;
dur_off=5*round(1200*durms_off/1000);
endOff=find( (fdhn_off) > (length(dhn)-dur_off) );
if isempty(endOff),
   [val_endOff,endOff]=max(fdhn);
end
```

```
endOff(1);
endOn=find(fdhn_off > dur_off);
prevOn=0;
prevOff=0;

e=1; % keep count of ons ;
d=1; % keep count of offs;

% identify plausible ons if duration ms is empty of spikes before, and full after
on_count=[];
w=startOn(1);
while (fdhn(w) < length(dhn)-dur-1) & (w < length(fdhn))
      issonbefore=[];
      issonafter =[];
      issonbefore = find(dhn(fdhn(w)-dur:fdhn(w)-1) ~= 0);  % must be empty
      issonafter  = find(dhn(fdhn(w)+1:fdhn(w)+dur) ~= 0);  % must be full

       if(isempty(issonbefore)==1 & isempty(issonafter)==0)
            on_count(e)=w; % <-- use for percent phase calc.
            on(fdhn(w))=max(eta(:,cn))/10;
            e=e+1;
         else
           on(fdhn(w))=0;
       end
       w=w+1;
end

% identify plausible offs if duration ms is full of spikes before, and empty after
off_count=[];
z=endOn(1);
while (fdhn_off(z) < length(dhn)-dur-1) & (z < length(fdhn_off))

   issoffbefore=[];
   issoffafter =[];

   issoffbefore = find(dhn(fdhn_off(z)-dur:fdhn_off(z)-1) ~=0 );  % must be full
   issoffafter  = find(dhn(fdhn_off(z)+1:fdhn_off(z)+dur) ~=0 );  % must be empty

   if (isempty(issoffbefore)==0 & isempty(issoffafter)==1)
      off_count(d)=z;
      off(fdhn_off(z))=max(eta(:,cn))/10;
      d=d+1;
   else
     off(fdhn_off(z))=0;
   end
z=z+1;
```

```
end

% User confirms identified sections of activation
figure('Position',[7 70 1010 612]), hold on
%plot(Sa), hold on,
plot(eta(:,cn),'k'), plot(on,'g'), plot(off,'r')
axis([0 LS -0.5 max(eta(:,cn))]);
range=get(gca,'YLim');
zline=zeros(1,LS);
hold on,plot(zline,'w')
title('Pick ON points (g) you want to KEEP, hit enter, then repeat for OFF points (r)')

% Perfect aim in picking points is not required. This finds the nearest peaks to the
% already determined array of peaks that satisfy the classification conditions
keepon=ginput;
keepon=keepon(:,1);
keepingon=nearest(fdhn(on_count),keepon)';

keepoff=ginput;
keepoff=keepoff(:,1);
keepingoff=nearest(fdhn_off(off_count),keepoff)';

% Plot the final peaks
figure('Position',[7 70 1010 612])
plot(Sa,'k'), hold on
[pospeakind,negpeakind]=peakdetect(eta(:,cn));
autoaddlines(keepingon,range(2),'g',1,'-');
peakon=nearest(pospeakind,keepingon)'
autoaddlines(keepingoff,range(2),'r',1,'-');
peakoff=nearest(pospeakind,keepingoff)'

figure,plot(Sa)
autoaddlines(peakon,range(2),'g',1,'-');
autoaddlines(peakoff,range(2),'r',1,'-');
title('If this does not look right, or you missed a point, hit ^c, and type NuOnDetect at the
prompt')

% dump to screen
disp('You have used the following input parameters:')
disp('   D  lo cf   hi cf   cn   ms')
if reconstruct == 1
   disp([ NoiseLevel   ArtifactLevel   NoiseLevelD cn  durms])
   variables=[NoiseLevel   ArtifactLevel   NoiseLevelD cn  durms];
else
   disp([NoiseLevel   1   maxscale   cn  durms])
   variables=[NoiseLevel   1   maxscale   cn  durms];
```

```
end

disp('resulting in the following output parameters:')
disp('  SNR      M       eta      th')
disp([SR   M   gammas(cn) th(cn)])

% build output
spon=length(peakon);
spoff=length(peakoff);
if ((spoff-spon)>0), padd=spoff-spon; paddz=zeros(1,padd)'; peakon=[peakon; paddz];,
end;
if ((spoff-spon)<0), padd=spon-spoff; paddz=zeros(1,padd)'; peakoff=[peakoff; paddz];,
end;
onoffpeaks=[peakon peakoff];

fid=fopen([subjlabel wave 'D' num2str(NoiseLevel) 'wavelet.txt'],'w');
fprintf(fid,'subject:\t %s\n',subjlabel);
fprintf(fid,'input vars\t wavelet type\t  Noise is level (D)\t lo cf level\t hi cf level\t
threshold iteration (cn)\t check dur (ms)\n');
fprintf(fid,'\t%s\t %2.0f\t %2.0f\t %2.0f\t %2.0f\t %2.1f\t\n',wave,variables);
fprintf(fid,'output vars\t SNR (dB)\t  Max Noise (M)\t eta\t  threshold\t\n');
fprintf(fid,'\t %3.2f\t  %5.3f\t %2.1f\t  %5.3f\t\n',SR,M,gammas(cn),th(cn));
fprintf(fid, 'on:\t  off:\t\n');
fprintf(fid, '%6.0f\t %6.0f\t\n',onoffpeaks');
fclose(fid);
```

REFERENCES

[54] Hodges, P., & Bui, B. (1996). A Comparison of Computer-Based Methods for the Determination of Onset of Muscle Contraction using Electromyography. Electroencephalography and Clinical Neurophysiology, 101, 511-519.

[55] Hsu, P.-H., & Tseng, Y.-H. Multiscale Analysis of Hyperspectral Data Using Wavelets for Spectral Feature Extraction . Hyperspectral & Data Acquisition Systems: Conference Proceedings of ACRS 2000, The 21st Asian Conference on Remote Sensing GISdevelopment.net.

[56] Ismail, A. R., & Asfour, S. S. (1998). Continuous wavelet transform application to EMG signals during human gait . Signals, Systems & Computers, Conference Record of the Thirty-Second Asilomar Conference on (pp. 325-329).

[57] Kang, W. J., Cheng, J. S., Shiu, J. R., & Kuo, T. S. (1996). A comparative analysis of various EMG pattern recognition methods. Med. Eng. Phys., 18(5), 390-395.

[58] Karlsson, S., & Gerdle, B. (2001). Mean frequency and signal amplitude of the surface EMG of the quadriceps muscles increase with increasing torque--a study using the continuous wavelet transform. J Electromyogr Kinesiol, 11(2), 131-40.

[59] Karst, G. M., & Hasan, Z. (1991). Initiation rules for planar, two-joint arm movements: agonist selection for movements throughout the work space. J Neurophysiol, 66(5), 1579-93.

[60] Karst, G. M., & Hasan, Z. (1991). Timing and magnitude of electromyographic activity for two-joint arm movements in different directions. J Neurophysiol, 66(5), 1594-604.

[61] Kashiwagi, K., Tanaka, M., Kawazoe, T., Furuichi, K., & Takada, H. (1995). Effect of Amplitude Normalization on Surface EMG Linear Envelopes of Masticatory Muscles during Gum Chewing. Journal of Osaka Dental University, 29(1), 19-28.

REFERENCES

[62] Kasman, G. S., Cram, J. R., & Wolf, S. L. (1998). Clinical applications in Surface Electromyography; Chronic Musculoskeletal Pain. Gaithersburg MA, USA: Aspen Publishers, Inc.

[63] Khalil, M., & Duchene, J. (2000). Uterine EMG analysis: a dynamic approach for change detection and classification. IEEE Trans Biomed Eng, 47(6), 748-56.

[64] Khalil, M., & Duchene, J. (2000). Uterine EMG analysis: a dynamic approach for change detection and classification. IEEE Trans Biomed Eng, 47(6), 748-56.

[65] Kim, K. H., & Kim, S. J. (2003). A Wavelet-based method for action potential detection from extracellular neural signal recording with low signal-to-noise ratio. IEEE Trans Biomed Engr, 50(8), 999-1011.

[66] Kohler, B. U., Hennig, C., & Orglmeister, R. (2002). The principles of software QRS detection. IEEE Eng Med Biol Mag, 21(1), 42-57.

[67] Kollmitzer, J., Ebenbichler, G., & Kopf, A. (1999). Reliability of Surface Electromyographic Measurements. Clinical Neurophysiology, 110, 725-734.

[68] Kreyzig, E. (1979). Advanced Engineering Mathematics. New York, USA: John Wiley & Sons.

[69] Laterza, F., & Olmo, G. (27 Feb. 1997). Analysis of EMG signals by means of the matched wavelet transform . Electronics Letters, 33(5), 357-359 .

[70] Leader, J. K., Boston, J. R., & Moore, C. A. (1998). A data dependent computer algorithm for the detection of muscle activity onset and offset from EMG recordings. Electroencephalogr Clin Neurophysiol, 109(2), 119-23.

REFERENCES

[71] Lee, S., Kim, J., & Park, S. (1996). An Enhanced Feature Extration Algorithm for EMG Pattern Classification. IEEE Transactions on Rehabilitation Engineering, 4(4), 439-443.

[72] Lehman, G., & McGill, S. M. (1999). The Importance of Normalization in the Interpretation of Surface Electromyography: A Proof of Principle. Journal of Manipulative and Physiological Therapeutics, 22(7), 444-446.

[73] Li, B.-L., & , W. H. Change detection in nonlinear biosignal processing by wavelets. Biomedical Engineering Conference, 1996., Proceedings of the 1996 Fifteenth Southern .

[74] Lidierth, M. (1986). A computer based method for automated measurement of the periods of muscular activity from an EMG and its application to locomotor EMGs. Electroencephalography and Clinical Neurophysiology, 64, 378-380.

[75] Malbos, F., Baudry, M., & Montresor, S. Detection of stop consonants with the wavelet transform. Time-Frequency and Time-Scale Analysis, 1994., Proceedings of the IEEE-SP International Symposium on (pp. 612-615).

[76] Mambrito, B., & De Luca, C. J. (1984). A technique for the detection, decomposition and analysis of the EMG signal. Electroencephalogr Clin Neurophysiol, 58(2), 175-88.

[77] Merlo, A., Farina, D., & Merletti, R. (2003). A fast and reliable technique for muscle activity detection from surface EMG signals. IEEE Trans Biomed Engr, 50(3), 316-323.

[78] Micera, S., Sabatini, A., & Dario, P. (1998). An algorithm for Detecting the Onset of Muscle Contraction by EMG Signal Processing. Medical Engineering & Physics, 20, 211-215.

[79] Mirka, G. (1991). The quantification of EMG normalization error. Ergonomics, 34(3), 343-352.

REFERENCES

[80] Misiti, M., Misiti, Y., Oppenheim, G., & Poggi, J.-M. (1996). Massachusetts: The MathWorks, Inc.

[81] Mochimaru, F., & Fujimoto, Y. (2002). Detecting the fetal electrocardiogram by wavelet theory-based methods. Progress in Biomedical Research, 7(3), 185-193.

[82] Moshou, D., Hostens, I., Papaioannou, G., & Ramon, H. (2000). Wavelets and self-organizing maps in electromyogram (EMG) analysis. ESIT 2000 .

[83] Netz, J., Lammers, T., & Homberg, V. (1997). Reorganization of motor output in the non-affected hemisphere after stroke. Brain, 120 (Pt 9), 1579-86.

[84] Ozdamar, O., & Kalayci, T. (1998). Detection of spikes with artificial neural networks using raw EEG. Comput Biomed Res, 31(2), 122-42.

[85] Page, S. J. (2000). Imagery improves motor function in chronic stroke patients with hemiplegia: A pilot study. Occ Ther J Res, 20(3), 200-15.

[86] Page, S. J., Levine, P., Sisto, S., & Johnston, M. V. (2001). A randomized efficacy and feasibility study of imagery in acute stroke. Clin Rehabil, 15(3), 233-40.

[87] Page, S. J., Levine, P., Sisto, S. A., & Johnston, M. V. (2001). Mental practice combined with physical practice for upper-limb motor deficit in subacute stroke. Phys Ther, 81(8), 1455-62.

[88] Panagiotacopulos, N. D., Lee, J. S., Pope, M. H., & Friesen, K. (1998). Evaluation of EMG signals from rehabilitated patients with lower back pain using wavelets. J Electromyogr Kinesiol, 8(4), 269-78.

[89] Panagiotacopulos, N. D., Lee, J. S., Pope, M. H., Magnusson, M. L., Wilder, D. G., Friesen, K., & Stielau, W. (1997). Detection of wire EMG activity in whiplash injuries using wavelets. Iowa Orthop J, 17, 134-8.

REFERENCES

[90] Pattichis, C. S., & Pattichis, M. S. (1999). Time-scale analysis of motor unit action potentials. IEEE Trans Biomed Eng, 46(11), 1320-9.

[91] Perry, J., Bontrager, E. L., Bogey, R. A., Gronley, J. K., & Barnes, L. A. (1993). The Rancho EMG analyzer: a computerized system for gait analysis. J Biomed Eng, 15(6), 487-96.

[92] Perry, J. (1992). Gait Analysis, Normal and Pathological Function. Dynamic Electromyography (pp. 381-411). New York, NY, USA: McGraw Hill.

[93] Petrock, A., Reisman, S., Weiner, S., & Siegel, A. (2002). Wavelet analysis of messeter muscle activity resulting from stimulation of hypothalamic behavioral sites. Proceedings of the IEEE 28th annual northeast Bioengineering conference, 2002. (pp. 49-50). IEEE .

[94] Petrock, A. M. (2002). Masseter Muscle Activity Resulting From Stimulation Of Hypothalamic Behavioral Sites: A Wavelet Analysis. Unpublished doctoral dissertation, New Jersey Institute of Technology, Newark, NJ.

[95] Pfeiffer, G., & Kunze, K. (1992). Turn and phase counts of individual motor unit potentials: correlation and reliability. Electroencephalography and Clinical Neurophysiology, 85, 161-165.

[96] Pope, M. H., Aleksiev, A., Panagiotacopulos, N. D., Lee, J. S., Wilder, D. G., Friesen, K., Stielau, W., & Goel, V. K. (2000). Evaluation of low back muscle surface EMG signals using wavelets. Clin Biomech (Bristol, Avon), 15(8), 567-73.

[97] Provaznik, I., & Kozumplik, J. Wavelet-based compression of ECG signals. Proceedings of the 18th Annual International Conference of the IEEE, Bridging Disciplines for Biomedicine. (pp. 1210-1211).

[98] Provaznik, I. (2002). Wavelet Analysis for signal detection - applications to experimental cardiology research. Unpublished doctoral dissertation, Faculty o fElectrical Engineering and Communication, Department of Biomedical engineering, Brno University of Technology, Brno.

REFERENCES

[99] Rash, G. (1985) Electromgraphy Fundamentals, GCMAS Instructional Documents [Web Page]. URL http://www.gcmas.org/basic_documents.html.

[100] Reza, A. M. (1999) From Fourier Transform to Wavelet Transform; Basic Concepts [Web Page]. URL www.xilinx.com/products/logicore/dsp/fft_to_wavelet.pdf.

[101] Rioul, O., & Vetterli, M. (1991). Wavelets and signal processing. Signal Processing Magazine, IEEE, 84, 14-38.

[102] Sherwood, A. M., McKay, W. B., & Dimitrijevic, M. R. (1996). Motor control after spinal cord injury: assessment using surface EMG. Muscle Nerve, 19(8), 966-79.

[103] Sloboda, W. M., & Zatsiorsky, V. M. Wavelet Analysis of EMG Signals. Twenty First Annual Meeting of the American Society of Biomechanics .

[104] St-Amant, Y., Rancourt, D., & Clancy, E. (1998). Influence of Smoothing Window Length on Electromyogram Amplitude Estimates. IEEE Transactions on Biomedical Engineering, 45(6), 795-800.

[105] Stashuk, D. W. (1999). Decomposition and quantitative analysis of clinical electromyographic signals. Med Eng Phys, 21(6-7), 389-404.

[106] Staude, G., & Wolf, W. (1999). Objective motor response onset detection in surface myoelectric signals. Med Eng Phys, 21(6-7), 449-67.

[107] Staude, G. H. (2001). Precise onset detection of human motor responses using a whitening filter and the log-likelihood-ratio test. IEEE Trans Biomed Eng, 48(11), 1292-305.

[108] Stulen, F. B., & DeLuca, C. J. (1981). Frequency parameters of the myoelectric signal as a measure of muscle conduction velocity. IEEE Trans Biomed Eng, 28(7), 515-23.

REFERENCES

[109] Stéphane Mallat. (1998). A Wavelet Tour of Signal ProcessingAcademic Press.

[110] Takada, K., Nagata, S., Miyawaki, R., Kuriyama, Y., & Yasuda, S. M. (1992). Automatic detection and measurement of EMG silent periods in masticatory muscles during chewing in man . Electromyogr. Clin. Neuropsysiol, 32, 499-505.

[111] Thilmann, A. F., Fellows, S. J., & Garms, E. (1990). Pathological stretch reflexes on the "good" side of hemiparetic patients. J Neurol Neurosurg Psychiatry, 53(3), 208-14.

[112] Trombly, C. A. (1992). Deficits of reaching in subjects with left hemiparesis: a pilot study. Am J Occup Ther, 46(10), 887-97.

[113] Unser, M., & Aldroubi, A. (1996). A Review of Wavelets in Biomedical Applications. Proceedings of the IEEE, 84(4), 626-638.

[114] Valens, C. (1999) A Really Friendly Guide to Wavelets [Web Page]. URL http://perso.wanadoo.fr/polyvalens/clemens/wavelets/wavelets.html.

[115] Valens, C. (Really Friendly Wavelet Guide <crudesoft@yahoo.com>). (2003, October 31). Re: wavelet tutorial and question. E-mail to Janina Wilen (jwilen@kmrrec.org).

[116] Van Boxtel, G. J. M., Geraats, L. H. D., Van Den Berg-Lenssen, M. M. C., & Brunia, C. H. M. (1993). Detection of EMG onset in ERP research. Psychophysiology, 30, 405-412.

[117] Walter, C. B. (1984). Temporal Quantification of Electromyography with Reference to Motor Control Research. Human Movement Science, 3, 155-162.

[118] Weiss, T., Hansen, E., Beyer, L., Conradi, M. L., Merten, F., Nichelmann, C., Rost, R., & Zippel, C. (1994). Activation processes during mental practice in stroke patients. Int J Psychophysiol, 17(1), 91-100.

REFERENCES

[119] Whitall, J., McCombe Waller, S., Silver, K. H., & Macko, R. F. (2000). Repetitive bilateral arm training with rhythmic auditory cueing improves motor function in chronic hemiparetic stroke. Stroke, 31(10), 2390-5.

[120] Wilen, J., Sisto, S., & Kirshblum, S. (2001). Reliability of an Algorithm for the Detection of Muscle Activation in Surface Electromyograms During Periodic Wheelchair Propulsion. Gait and Posture: Gait and Clinical Movement Analysis Society Meeting (p. 283).

[121] Wilen, J., Sisto, S. A., & Kirshblum, S. (2002). Algorithm for the detection of muscle activation in surface electromyograms during periodic motion. Annals of Biomedical Engineering, 30(1), 97-106.

[122] Wilson, L. R., Gandevia, S. C., Inglis, J. T., Gracies, J., & Burke, D. (1999). Muscle spindle activity in the affected upper limb after a unilateral stroke. Brain, 122 (Pt 11), 2079-88.

[123] Winter, D. A. (1990). Biomechanics and Motor Control of Human Movement (2 ed.).Wiley, John & Sons, Inc.

[124] Witte, H., Eiselt, M., Patakova, I., Petranek, S., Griessbach, G., Krajca, V., & Rother, M. (1991). Use of discrete Hilbert transformation for automatic spike mapping: a methodological investigation. Med Biol Eng Comput, 29(3), 242-8.

[125] Yaguez, L., Nagel, D., Hoffman, H., Canavan, A. G., Wist, E., & Homberg, V. (1998). A mental route to motor learning: improving trajectorial kinematics through imagery training. Behav Brain Res, 90(1), 95-106.

[126] Zennaro, D., Wellig, P., Koch, V. M., Moschytz, G. S., & Laubli, T. (2003). A software package for the decomposition of long-term multichannel EMG signals using wavelet coefficients. Biomedical Engineering, IEEE Transactions on, Vol.50, (Iss.1), 58-69.

REFERENCES

[127] Zhou, P., Rymer, W. Z., Suresh, N., & Zhang, L. (2001). A study of surface motor unit action potentials in first dorsal interosseus (FDI) muscle . Engineering in Medicine and Biology Society, 2001. Proceedings of the 23rd Annual International Conference of the IEEE, 2, 1074-1077 .