FUNDAMENTALS ELECTROMYOGRAPHY (EMG)

Introduction

Electromyography (EMG) is a technique used for evaluating the physiological, biochemical, and electrical activity of skeletal muscles during contraction. The EMG signal reflects the effect of an external excitation (or pathological condition) on a skeletal muscle. This effect is represented by a time evolving plot. In this plot, the horizontal axis is used to represent the evolving time in milliseconds (mesa) and the vertical axis represents the magnitude of electrical potential of the effect of this excitation in mill volts (mV). This electrical potential is generated by muscle cells when these cells are activated. The EMG signal, if it is properly analyzed, it will reflect the **physiological properties of the skeletal muscle.** That is, contractibility (ability of muscle cells to forcefully shorten), excitability (ability of muscle cells to forcefully activated), extensibility, and elasticity. Furthermore, it will provide information about **muscle fatigue** (the inability of the muscle to do work) which is probably the most important physiological property of a muscle condition.

The EMG signal analysis could also help to: detect medical abnormalities, activation level, recruitment order, normal/abnormal neuromuscular functioning, and to analyze the biomechanics of human movement.

Note that here we are referring to data obtained from surface EMG (S-EMG) measurements.

There are two technique used for the processing of S-EMG signals: (1) The time domain and (2) the frequency domain. In the case of time domain, researchers applied techniques such as integration, linear envelope, root-mean-square (RMS), and count zero-crossings, while in the case of frequency domain**,** Fourier transformations (FFT) have been used to transform a time domain signal into the frequency domain.

FFT has been used to obtain the frequency spectrum of the S-EMG signal which was recorded during sustained (isometric) muscle contraction, which in turn has been used to detect muscle fatigue, force production, and muscle fiber signal conduction velocity. In addition, the study of the power spectrum can provide information about spatial and temporal recruitment of motor units. It has been suggested that each muscle may have a characteristic power spectrum.

Mean and median frequencies are also measured by FFT and shifts to lower frequencies during sustained muscle contraction have been reported in. However, there are some controversies regarding the shift of median and mean frequencies as a result of muscle fatigue. Some studies report that the median frequency is a more reproducible measure of muscle fatigue than the mean frequency because it is less sensitive to noise and perturbation of the signal. Other studies have shown that mean frequency, rather than median frequency, yields a more sensitive measure of spectral shifts. The shifts to lower frequencies of the median and/or mean frequency usually corresponds to physiochemical changes associated with fatigue, changes in the action potential transmission velocity, the characteristics of the types of motor units (fast or slow), and their ratio in a particular muscle or muscle group.

It should be pointed out here that FFT is a suitable method for stationary signals. However, S-EMG is a non-stationary signal (signal with time varying frequency), and for that reason in this study, we are using a powerful mathematical method which is suitable for non-stationary signals. This method is known as Wavelet Transform (WT) and enables us to obtain more information about the S-EMG signal we are dealing with. More specifically, WT analysis will help us to view the signal at various scales and thus enable us to obtain a far more detailed frequency analysis and identify easily signal features difficult to recognize otherwise. Furthermore, WT it is a more powerful tool than Fourier methods for S-EMG analysis, since it provides an optimum time-frequency resolution.

In this study, there are two types of skeletal muscle fibers that we are interested in: **slow-twitch** (type I) and **fast-twitch** (type II). **Type I** are **slow twitch** fibers with **high endurance** (long-endurance such as distance running,), while **Type II** are **fast twitch** fibers with lower endurance (these muscles fatigue faster but are used in powerful bursts of movements like sprinting).

1. The Central Nervous System (CNS)

Figure 1 shows a simplified functional diagram of the **central nervous system** and how an EMG signal is acquired.
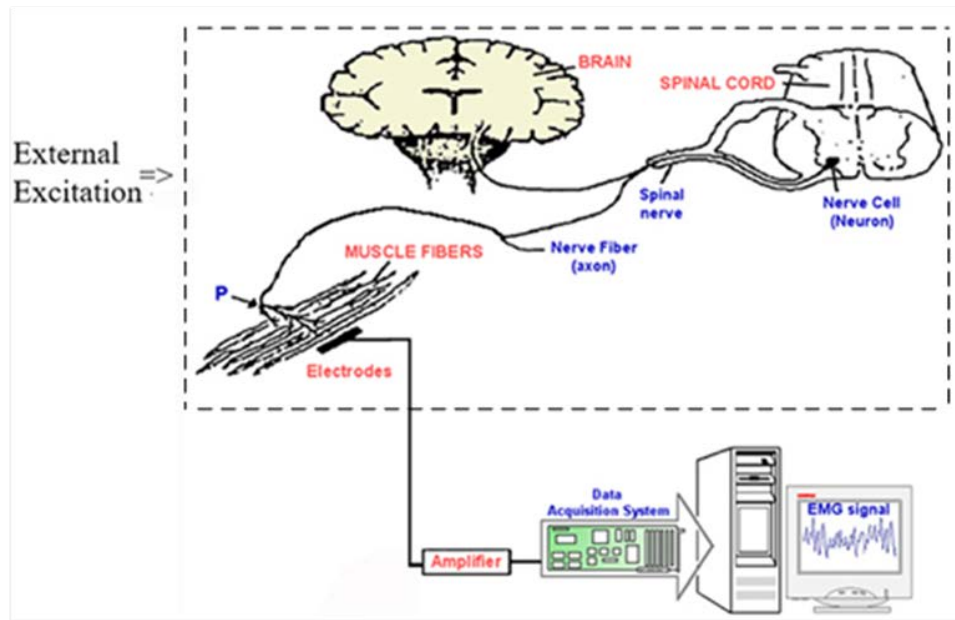
Fig 1. Central Nervous System and a typical EMG acquisition system.

2. Anatomical structure of Skeletal Muscle

The muscle is surrounded by a connective tissue called **epimysium** and is divided by fascicles by **perimysium** which contain several **muscle fibers**. The structural unit of skeletal muscle is the muscle fiber, or cell (Fig. 2). A muscle cell is a thin structure ranging from 10 to 100 microns in diameter and from a few millimeters to 30 cm in length. The muscle fibers do not extend the entire length of the muscle. Instead, the cells are attached to either the origin or insertion tendon at one end and connective tissue septa at the other end. The muscle fiber is further subdivided into **myofibrils** where thick and thin filaments are arranged longitudinally. **A muscle fiber produces force by contracting.**
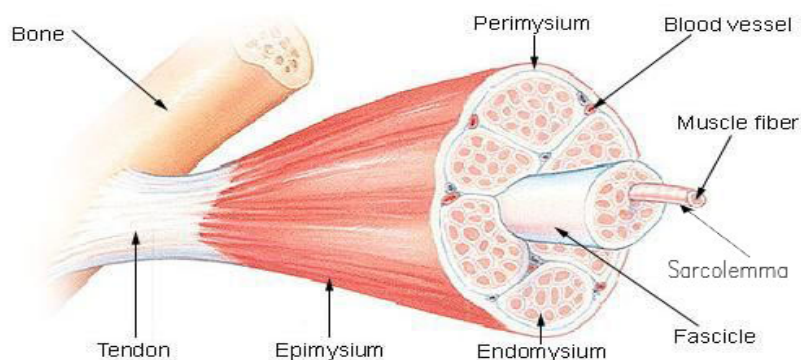

Fig 2. The organization of skeletal muscle

A muscle fiber is surrounded by the **sarcolemma**. The sarcolemma is a thin semi permeable membrane composed of a lipid bilayer that has channels by which certain ions can move between the intracellular and the extra cellular fluid.

Muscle fibers are always grouped together sharing the same nerve fiber. This nerve fiber (**axon**), refer to Fig 1 is the transmission line of a nerve cell and runs from the spinal cord to the peripheral muscle. A nerve cell which innervates a group of muscle fibers is known as **motor-neuron**. Every cell is located close to the spinal cord.

Muscle fatigue is the inability of a muscle to generate force. It can be a result of vigorous exercise, but abnormal fatigue may be caused by barriers to or interference with the different stages of muscle contraction.

3. **Action Potential (AP)**

There is a difference in the composition of the extra cellular fluid and intra cellular fluid in resting state. The resting membrane potential is the voltage difference across the plasma membrane induced by the electrochemical potential difference. The major ions that maintain the resting potential are Na^+ K^+, Cl^-, etc. When substantially large stimulus is applied, an action potential is triggered. This takes place only when the depolarization is sufficient to reach a threshold value (Fig 3).

Electrical muscle activity, **electromyography (EMG)**, is the result of an external excitation of the muscle fiber (cell) which usually runs along the whole muscle from tendon to tendon. In its resting state, there is a potential of approximately -90mV across the cell membrane (the fiber wall) with zero reference on the outside. The fiber is an 'on-off device' in the sense that mechanical work is produced in twitches only, which have durations ranging from 35 to 75 ms (1).
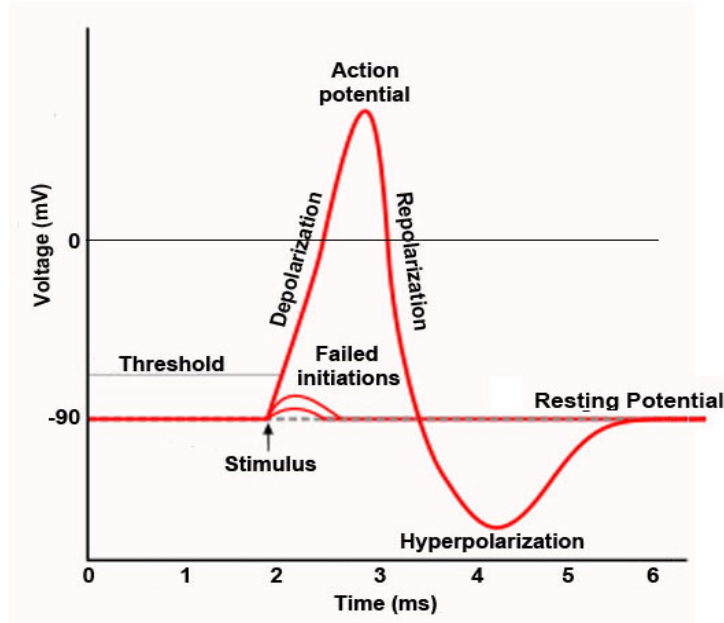
Fig 3. Action potential

An **action potential** is a rapid change in the membrane potential followed by a return to the resting membrane potential. An action potential is propagated with the same shape and size along the whole length of a nerve or muscle cell.

4. **Discussion on Motor Unit (MU) Recruitment Process (Activation)**

The smallest functional unit of muscle contraction is the **motor unit** (Figs 1 and 4). A motor unit consists of a group of muscle fibers innervate by a single **motor neuron** (in other words, a **motor unit** consists of one motor neuron and all the muscle fibers it stimulates). Within a single motor unit, individual muscle fibers discharge nearly at the same time.
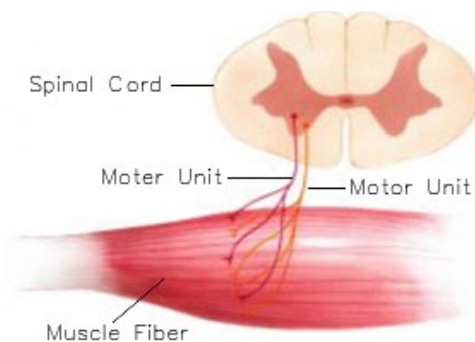

Fig 4. Motor Unit

5

Motor unit recruitment refers to the activation of additional motor units to accomplish an increase in **contractile strength** in a muscle. The higher the recruitment the stronger the muscle contraction will be. In physiology, an action potential occurs when the membrane potential of a specific cell location rapidly rises and falls: this depolarization then causes adjacent locations to similarly depolarize.

When an action potential is evoked in a motor nerve by a critical depolarization, an endplate, or the neuromuscular junction, potential is produced by **acetylcholine** release. Acetylcholine stimulates the membrane of the muscle fiber, or sarcolemma. This triggers rapid depolarization (about +20mV) and repolarization of the muscle fiber (Fig.3). The action potential is propagated along the sarcolemma and into the muscle fiber through the transverse tubules. The generation of an action potential takes place in all the muscle cells in the motor unit and is followed by synchronous contractions of all the muscle cells in the motor unit. A motor-neuron together with its muscle fibers is called **motor unit (MU)**. The number of muscle fibers coupled in one MU range from 3 (in small muscles) up to 1000 (in large muscles). All fibers belonging to the same MU are triggered simultaneously. **The AP seen in by an electrode recording is the superposition of all the APs of the muscle fibers belonging to the triggered MU (MU-AP).** The MU is usually activated repeatedly with a firing frequency of 10-50 Hz. A whole muscle usually consists of several hundreds of MUs.

The **force development** of the muscle is modulated by the CNS by **recruitment** of an appropriate number of MUs and by varying the firing frequencies of the recruited MUs. The MUs seem to be recruited in a fixed order according to the reversed size order of their motor-neuron.

The depolarization generates an electric field near the muscle fibers which can be detected by a skin surface electrode located near this field. The resulting signal is called the muscle fiber action potential. The combination of the muscle fiber action potential from all the muscle fibers of a single motor unit is the **motor unit action potential**.

When an electrical impulse originated from the central nervous system (CNS) reaches a point on the muscle (a point at which a nerve fiber comes in contact with a muscle fiber called **innervation point**) of the muscle fiber, a twitch is triggered via a biochemical transmission process. The biochemical process of transferring **chemical energy** into **mechanical energy** is elicited on both sides of the innervation point and spreads in both directions towards the ends of the muscle fiber. The biochemical process is accompanied by a transient drop of the membrane potential called **depolarization**. The electrical activity of the depolarization is detected with an electrode in the vicinity of the fiber. This signal is referred to as an action potential (AP). The AP resembles the transient membrane potential in Fig. 3

5. **Surface Electrodes**

Surface EMG (S-EMG) is a non-invasive technique that measures signals containing certain temporal characteristics useful for understanding the muscle's response. Surface electrodes, place over the muscle, and register summated activities from many motor units. S- EMG can be obtained by using electrodes affixed to the surface of the skin. With the development of very sophisticated electrodes, S-EMG is now being used in many areas of ergonomics, sports medicine and even in clinical applications. S-EMG is a very useful tool in detecting integrated muscular behavior.

One **advantage** of the S-EMG technique is the convenience in terms of relative ease of application and lessened discomfort of the subject. On the other hand, there are few **disadvantages** in using the S-EMG technique. The surface electrodes do not have selectivity to a specific muscle fiber because of the wide pickup are from the muscle as compared to fine wire electrodes. The signals detected from the pickup area may originate in a deep muscle, or even worse, from different muscle groups.

Therefore, there are concerns about the validity of a recording. Although efforts have been made to quantify and determine the effect of cross talk, there is no established or easy way for S- EMG to eliminate this problem. In general, however, S-EMG is satisfactory for the analysis of temporal, force, or fatigue relationships. The main advantage of surface electrode can be most

effectively utilized when the simultaneous activity of many muscles is being studied in a large group of muscles.

In summary, the EMG signal can be detected either by intramuscular electrodes (needle or wire) or by surface electrodes attached to the skin over the studied muscle. In both cases the EMG is the sum of the contributions from a large number of MU-APs in the muscle. The MU-AP is attenuated substantially on its way through the tissue. Hence, with intramuscular electrodes, MU-APs adjacent to the electrode, strongly dominate over contributions from more distant MU-APs. With surface electrodes, a larger number of MUs contribute. Since the amplitude of the MU-AP rapidly decreases with distance, the main contribution to an S-EMG comes from the superficial part of the muscle. Hence, S-EMG is restricted to superficial muscles.

The electrically conducting part of a surface electrode usually consists of a silver-covered surface (10-50 mm2) which is taped or glued to the skin. The impedance between electrode and skin is reduced by applying some kind of electrode paste. The standard procedure is to use two electrodes, 10-20 mm apart, placed over the belly of the muscle and aligned in the muscle fiber direction. The two electrodes are connected to a differential amplifier with a third electrode placed somewhere else on the body as a reference. The input impedance and common mode rejection should be as high as possible.

6. Features of EMG

Considerable progress has been made when EMG analysis has been used in ergonomics to investigate topics like musculoskeletal injury, low-back pain, and muscle fatigue from overexertion. Although modern instrumentation has greatly facilitated the acquisition of EMG data, many issues remain unresolved in the interpretation of EMG signals.

The information obtainable from EMG analysis can be generally divided into the following three categories. (1) The relationship between temporal aspects of EMG and anatomically associated movement. (2) The relationship between EMG and production of force. (3) The relationship between EMG and muscle fatigue.

6.1 **Temporal Information**

The most basic information obtainable from an EMG recording is whether the **muscle** was **on** or **off** during an activity or at a particular time. In order for a muscle to be considered on, the EMG recording must exceed a certain threshold, whether defined by an arbitrarily of statistically predetermined level or by the noise level of the equipment responsible for the measurement. It often is more difficult to determine that a muscle is off because a muscle may infrequently be in a state of total relaxation.

Some researchers describe the threshold in an absolute value. This measure simply relates how active the muscle was during the experimental conditions. The measure is not an indication of muscle force, but simply a function of muscle usage. The signal can be quantified in several ways. Quantification may include peak activity, mean activity, activity as a function of a given position or posture, and rate of muscle activity onset.

To determine the on-off state, force, or fatigue present within a muscle, some form of EMG signal treatment usually is recommended and often required. If the raw or processed signal exhibited activity, the muscle was in use during the exertion. Differences do exist, however, between the temporal characteristics of the EMG and the produced tension. The most apparent is the pure time delay. Initial tension levels in the muscle also influence the delay times seen at onset.

Most studies that now investigate the on-off state of the muscle are interested in the phasing of the EMG activities under various experimental conditions. A quantitative evaluation of muscle on-off state was performed by Marras and Reilly, using statistical analysis of muscle event time's derived from processed EMG. They were interested in how the patterns of trunk muscle activation changed as the angular velocity of the trunk increased during controlled simulated lifting motion.

6.2 **EMG-Force Production**

An EMG-force measurement seeks to quantify the average number and firing rate of motor units contributing to a particular muscle contraction, and to relate the quantity to the actual force produced. With increasing intensity of a contraction, more and more units will be recruited, and the unit firing frequency will also be increased. The observed motor unit activity reflects these changes as the resulting interference pattern becomes denser and of greater amplitude. The signal can be processed to estimate a numerical value (usually a percentage of a maximum voluntary contraction) to the level of EMG activity associated with the generation of a corresponding force.

However, a great deal of confusion exists regarding the relationship between processed EMG and produced force. Researchers have long suggested that EMG could be used to represent the active control input of the muscle, and that some relationship must exist. Some researchers have presented EMG as a direct indication of muscle force while others have presented very complex models using these signals to predict force. Many researchers have attempted to investigate the force of a muscle by simply observing the rectified and averaged (in some cases integrated) EMG signal in terms of the absolute number of micro volts generated and associated with a particular activity. One aspect of this study investigated was whether or not the normalized surface EMG signal amplitude versus normalized force relationship was dependent on exercise level. The RMS value of the signal amplitude was used as the variant parameter because it is the parameter that more completely reflects the physiological correlates of the motor unit behavior during a muscle contraction. (This result showed almost linear relationship between normalized RMS and force).

6.3 **EMG-Muscle Fatigue**

Very little is known about general physical fatigue that is experienced after heavy work. Metabolic changes, such as an increased blood **lactate** concentration and a fall in pH, may contribute to the perception of fatigue but cannot fully explain the phenomenon.

Fatigue can be experienced slowly or rapidly depending on the type of work performed by a muscle. When the workload is relatively small, the slow motor units are initially excited. If the

task requires fast and forceful contractions, fast motor units are also recruited for maximum output. Rapid muscle fatigue can be observed in the latter case.

6.3.1 Localized Muscle Fatigue

We are interested in fatigue on a local level within the body. This is true particularly in musculoskeletal exertions. Typical externally visible symptoms of fatigue are loss of force production capabilities and localized discomfort and pain. This type of fatigue has become known as localized muscle fatigue (LMF). Muscle exertion levels do not necessarily need to be high to cause LMF. **Isometric contractions** of as low as 10% of maximum voluntary contraction have shown signs of LMF.

6.3.2 Spectral Charges in EMG during Fatigue

During LMF, changes occur in the surface recorded EMG signal. Two of the most commonly cited changes are a shift in the frequency content of the signal toward the low end and an increase in the amplitude. Many researchers have applied spectrum analysis methods to show the presence of fatigue. Lindstrom and DeLuca contend that the spectrum shift and amplitude increase is related. They state that tissue acts as low pass filter. As the frequency content of the original signal shifts to the lower frequencies, more energy is transferred through the tissues to the electrodes. This energy transfer, in turn, increases the amplitude of the recorded signal.

A group of investigators have demonstrated a decrease of power density in the high frequency region of the EMG signal and increase in the low frequency region during fatigue contractions. Many researchers have proposed physiologic explanations for the changes in amplitude and spectral characteristic. Lindstrom et al. has demonstrated that the frequency shifts were almost entirely dependent on the propagation velocity of the action potentials. The reduced propagation velocities have been linked to the production and accumulation of acid metabolites.

Some investigators, however, showed that the use of spectral analysis does not provide significant differences with fatigue. They may be related to the fact that the frequency of the EMG signal can be altered by many factors, for example, by the change in load or by the change in muscle length.

11

6.3.3 Muscle Fatigue through microvolt

As described above, the EMG signal will increase its amplitude while the muscle is exerting a given amount of force in an isometric contraction. This may be due to the need to recruit more motor units to perform the same amount of work as the muscle fiber fatigue. Thus, by observing the processed EMG signal of a given portion of the muscle under constant force conditions, a quantitative indicator of the degree of muscle fatigue can be established. It is also important to note that this trend is evident only with surface electrodes.

Limitations in Fatigue Analysis

Although the use of EMG in measuring localized muscle fatigue is well established and frequently used, the technique has limitations. It is important to understand some of these limitations before undertaking an EMG analysis in the field of ergonomics or biomechanics. The first problem is in the basis definition of fatigue. Because there is no universal definition of fatigue, agreement on the validity and meaning of EMG measures will be questioned. Other factors in LMF such as pain tolerance, motivation, and synergistic accommodation are not included in EMG analysis and yet have been argued to be important. Additionally, spectral shifts have been used for short-term contraction fatigue, but the use of EMG in long-term fatigue is questioned. For muscle fatigue that occurs over a longer period of time, possibly hours, the use of EMG has not been well established. In addition, shifts in the various EMG indices have been shown to decrease rapidly during the initial stages of a contraction, but do not decay as rapidly toward the end of a long session of work.

7. Processing of EMG

EMG signals are neither periodic nor deterministic. In other words, the statistical behavior of the EMG signal is not exactly the same for every time arbitrarily chosen. Indeed, they do not repeat with a definite time interval, and a single mathematical expression cannot specify a detected EMG signal for all time.

7.1 Time Domain

The raw (unprocessed) EMG signal reveals the myoelectric activity from a muscle. Generally, as a first step, visual inspection of the EMG is made to compare the signal information and the general character of the movement, and the valuable feature is determined. The rapid random fluctuations in the signal are ignored as being due to the random summation and subtraction of the many muscle fiber action potentials detected. Instead, attention is paid to the boundary or envelop of the EMG signal. This signal processing is context specific, intuitive and has a quantitative basis that many signal processing methods seeks to mimic and even exploit.

The interpretation of EMG has long been performed by visual inspection of the raw signal. The observer was able to identify from the raw signal when a muscle shows activity or not. The on-off, or active-active, temporal information from raw EMG signals, however, may sometimes cause difficulties in extracting interpretations of the muscle's behavior. According to the study of Winter, the on-off information can be different if a different threshold level is used. It is very likely that the result can be misrepresented.

The simple comparisons of peak-to-peak amplitudes are inaccurate since the raw EMG signal is highly complex. Even if the raw signal is recorded with high quality instruments, proper processing of the EMG is required for a correct interpretation. A few time-domain based EMG signal processing techniques are used for extracting useful information relevant to the purpose of the ergonomic or kinesiology study. Most of the processing techniques are used to detect the muscular activities.

7.2 Rectification

The raw EMG detected by surface electrodes and amplified by a linear differential amplifier is a bipolar signal and oscillates randomly with a zero-mean value. Rectification is one technique frequently used in EMG-processor designs to translate the raw signal to a single polarity.

$$\text{RECT } [x\,(t)] = |x\,(t)| \tag{1}$$

In most cases, full-wave rectification is preferred because it retains all the energy of the signal

7.3 Linear Envelope

The rectified signal still expresses the random nature of the amplitude of the signal. A useful approach for extracting amplitude-related information from the signal is to smooth the rectified signal. The concept of smoothing involves the suppression of the high-frequency fluctuation from a signal so that its deflections appear smoother. This may be recognized as low-pass filtering procedure. The smaller the bandwidth is, the greater the smoothing will be.

$$\overline{|x(t)|}_{t2-t1} = \frac{1}{t_2-t_1} \int_{t1}^{t2} |x(t)| \, dt \tag{2}$$

The equivalent operation to smoothing, in a digital sense, is averaging. The time window ($T = t_2 - t_1$) can be set for a certain interval and the move the window along the window; this is called moving average. This operation introduces a lag.

$$\overline{|x(t)|} = \frac{1}{T} \int_{t}^{t+T} |x(t)| \, dt \tag{3}$$

The output of the linear envelope detector represents a moving average of EMG activity. An undesirable side effect of the low-pass filter is the phase lag it causes in the envelope response. This lag may introduce significant errors in the measurement of temporally related variables. The phase lag can be avoided by using advanced signal processing techniques.

7.4 Integration

The most commonly used and abused data reduction procedure in EMG is the concept of integration. Integration refers to the mathematical operation of computing the area under the curve and the units are V·s or mV·ms. It is necessary to full wave rectify the raw signal to obtain the absolute value since the integral of the raw EMG is zero. The integral will increase continuously as a function of time. Usually the signal is integrated over a small area and then that is repeated over for entire signal.

$$I[x(T)] = \int_{0}^{t} |x(t)| \, dt \tag{4}$$

Typically integrating over fixed time interval is used. The integrated rectified value may be more useful, thereby indicating any time-dependent modifications of the signal.

$$I[\,x\,(T)] = \int_0^{t+T} |\,x\,(t)|dt \qquad (5)$$

Integrated EMG may represent the number of active motor units since it is related with the amplitude, duration and frequency of the action potentials. It can also provide temporal information in the form of relative muscle activity over a period.

7.5 RMS (Root-Mean-Square)

RMS processing is a method that allows consistent, valid, and accurate measurements of noisy, non-periodic, non-sinusoidal signals. Deluca and Van Dyke have demonstrated that the RMS value contains more relevant information than the mean rectified or integrated EMG.

$$RMS\,[\,x\,(t)] = \sqrt{\frac{1}{T}\int_0^{t+T} |\,x^2(t)dt} \qquad (6)$$

The RMS is not affected by the cancellation caused by the superposition of motor unit action potential trains. The RMS value is recommended more than previously described parameters.

7.6 Zero Crossings and Turns Counting

The zero-crossing method consists of counting the number of occurrences per unit time that the amplitude of the signal contains either a peak or crosses the zero value of the signal. This technique is not recommended for measuring the behavior of the signal as function of force or as a function of time during a sustained contraction. The turns or zeros and umber of motor unit action potential trains are linear for low level contractions. But as the contraction level increases, the additionally recruited motor units contribute motor unit action potential trains to the EMG signal. When the signal amplitude attains the characteristics of Gaussian random noise, the linear proportionality no longer holds.

The temporal aspects of the EMG-force relationship are affected by the specific processing methodology applied. The primary factor is the low-pass properties associated with the filtering function used. In most cases, such as integration and RMS, the filtering function is an exponential window with some associated time constant. A longer time constant produces a smoother estimate of the electrical activity of the muscle. This beneficial during static exertions where the electrical state of the muscle is stationary. During dynamic exertions, however, the response of the filter may be too slow to capture the changes occurring in the electrical state. The choice of an appropriate time constant is important for the type of activities under investigation.

The two most common techniques are integration and RMS, which have been shown to affect the temporal aspects of EMG differently. Additionally, the two common methods of estimating the RMS have different rise and full times. Thus, not only the time constant but also the dynamic aspects of the specific processor are important considerations if temporal information is to be derived from processed EMG.

7.7 Frequency Domain

Frequency domain processing is used to shift the electromyography's reference to the information content of the EMG signal from the time domain to the frequency domain. The value of this technique is in simplifying the identification and quantification of EMG information that manifests itself as changes in EMG frequency content. A normal use of this technique is to identify EMG frequency spectrum shifts believed to be related to localized muscle fatigue.

Fourier Transform and Power Spectrum

Fourier analysis has provided a great deal of applications to the signal analysis and processing area of EMG. The FFT is the most common method for determining the frequency spectrum of surface EMG signals recorded during sustained (isometric) muscle contraction. The power spectral density is the function commonly used for frequency domain analysis of EMG. It is defined as the Fourier transform of the autocorrelation function. An integration of the power spectral density over all frequencies yields the total power, hence the power spectrum.

16

The Fourier transform of a signal can be defined by

$$S(\omega) = \int_{-\infty}^{\infty} s(t)e^{-i\omega t}\, dt \tag{7}$$

The autocorrelation function, $\phi_{xx}(t)$, can also be defined by

$$\phi_{xx}(\tau) = \frac{1}{T}\int_0^T s(t)s(t-\tau)dt \tag{8}$$

Now, the power spectrum of a signal can be obtained by taking the Fourier transform of the autocorrelation function.

$$\phi_{xx}(\omega) = \int_{-\infty}^{\infty}\phi_{xx}(\tau)e^{-i\omega t}d\tau \tag{9}$$

The power spectrum represents the power density at frequency ω and measures the power content of the signal at each frequency. The area under the power spectral magnitude curve is equal to the power of the signal.

$$\phi_{xx}(\omega) = S(\omega)S(\omega) = |S(\omega)|^2 \tag{10}$$

Generally, two parameters of power density spectrum are used to access useful measures of the spectrum. The mean frequency is the average of all frequencies and the median frequency is the frequency that has equal distribution on each side. The mean and median frequencies are defined by the following equation:

$$S_{mean} = \frac{\int_0^{S_{med}} \omega\phi_{xx}(\omega)d\omega}{\int_{S_{med}}^S \phi_{xx}(\omega)d\omega} \tag{11}$$

$$\int_0^{S_{med}} \phi_x(\omega)\, d\omega = \int_{S_{med}}^{S_0} \phi_{xx}(\omega)\, d\omega \tag{12}$$

Of these two parameters, the median frequency was found to be less sensitive to noise. Fig 6 illustrates both mean and median frequency using an idealized fashion.

8. Motor Control and Measurement

8.1 Conscious Control

The movement behavior of the human body is often modeled by a closed-loop system. It is an effective model for understanding continuous, long-duration skills such as tracking, balancing, and slow positioning. Once a movement is triggered internally, its control is passed in an all-or-none fashion to the motor program for execution. It usually requires approximately 150 to 200 ms for modification once the movement is triggered.

8.2 Motor Reflexes

Reflexes can be defined as both involuntary and rapid responses to stimuli. A reflex is the modification in the relatively low-level processing in the spinal cord and brain stem, and often does not involve conscious control. The processes that contribute to the reflex can be studied by the EMG signal detected from a muscle that contributed to the reflex. An EMG signal detected from a muscle that contributed to the reflex. An EMG record from a muscle is shown in figure 7. The sequence of M1 responses, M2 responses, triggered reaction (not shown), and M3 responses are shown.
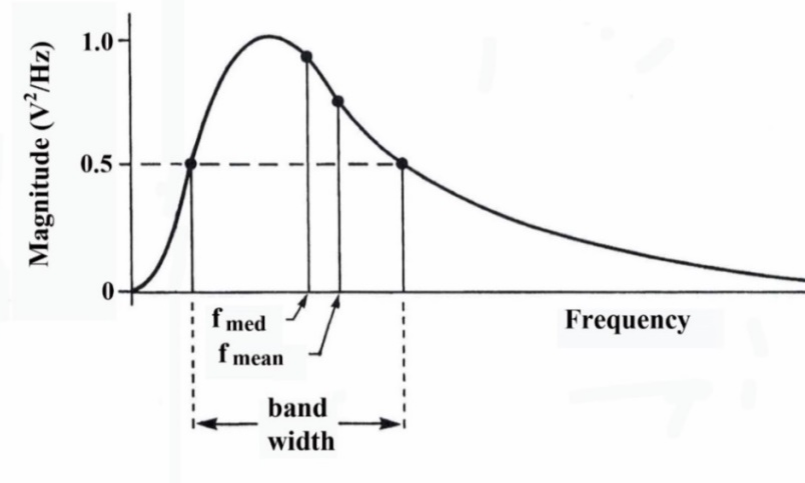

Fig 6. An idealized version of the frequency spectrum of the EMG signals

a. M1 Response

The most rapid and brief burst of EMG activity occurs at about 30 to 50 ms after the sudden load and is called the M1 response, M1 reflex, or monosynaptic stretch reflex. The M1 response is caused by the muscle spindles in the muscle being stretched when the load is added, which results in sensory information being sent to the spinal cord. After traveling to a single synapse in

18

the spinal cord, this information is routed directly back to the same muscle that was stretched, causing the increased contraction seen as the small EMG burst, The latency of this correction is very short because the information involves only one synapse and has a relatively short distance to travel.

b.   M2 Response

The second burst of EMG activity occurs at about 50 to 80 ms and is referred as the M2 response, M2 reflex, functional stretch reflex, or long-loop reflex. The M2 response generates more EMG activity than the M1 reflex and has a longer duration. Thus, it contributes far more to movement than the M1. This response also arises from the muscle spindles and travels to the spinal cord, but then the impulses go up the cord to higher centers in the brain (the motor cortex and / or the cerebellum). The longer travel distance and the additional synapses at the higher levels account for the longer delay in M2.

The M2 is more flexible than M1, allowing for a few other sources of sensory information to be integrated in the response. The M2 could be almost completely abolished and the M1 response world remains almost unmodified. The amplitude of the M2 response for a given input can be adjusted voluntarily to generate a powerful response when the goal is to hold the position as firmly as possible, or it can result in almost no response if the movement goal is to release under the increased load.

c.   Triggered Reaction

A third type of response, somewhat longer in latency than the M2, has been termed a triggered reaction. This action is also too fast to be a voluntary reaction, with latencies of 80 to 120 ms, but it is too slow to be an M2 response. It can affect musculature that is quite remote from the actual stimulation site, and is sensitive to the number of stimulus alternatives, similar to the reaction-time response which apparently be learned.
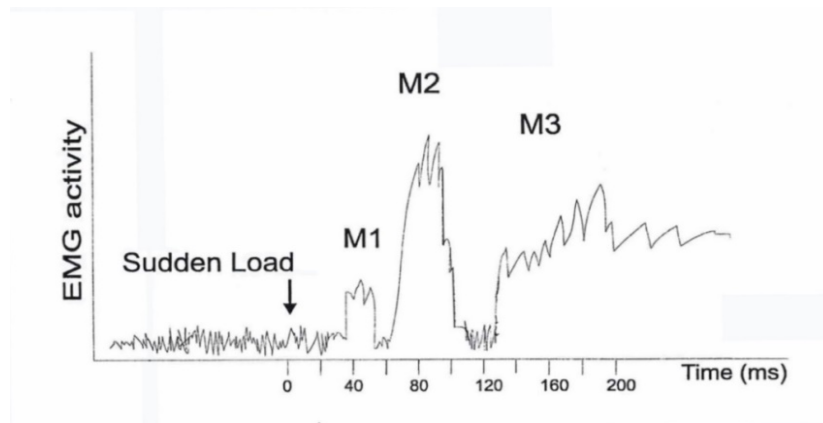
Fig 7. to a sudden EMG responses load

d. M3 Response

A final type of response to the added load is a voluntary reaction, sometimes called the M3 response. It is powerful and sustained, bringing the posture to the final position and holding it there. The latency of the M3 response is around 120 to 180 ms, depending on the task, and it can affect all the musculature, not just those muscles that are stretched. The M3 response is the most flexible of all, being modified by a host of factors such as instructions, anticipation, and so on.

9. Measurement of Time

The measurement has been used a great deal in motor behavior research. Reaction time and movement time are the most common parameters used in many research areas.

9.1 Research Time (RI)

Reaction time is interval of the time from the arrival of a sudden load to the beginning of response to it. In Fig.8, the subject is given a warning signal, and after a short fore period, the stimulus is presented. Thus, temporal anticipation, when the stimulus will arrive, was prevented while spatial anticipation, or which response to make is preserved.

During a substantial part of RT, the EMG is silent; indicating that the command to contract the muscle has not yet reached the muscle. Then, late in the response period in RT, the muscle is activated, but no movement occurs for a period. The interval from the signal to the first change in EMG is termed premotor RT and is thought to represent reflex and central processing. The

20

interval from the first change in the EMG signal to movement is termed motor RT and is presented as processes associated with the musculature itself.

9.2 Movement Time (MT)

Movement time is usually defined as the interval from the initiation of the response (the end of RT) to the completion of the movement. The sum of RT and MT is termed the response time.
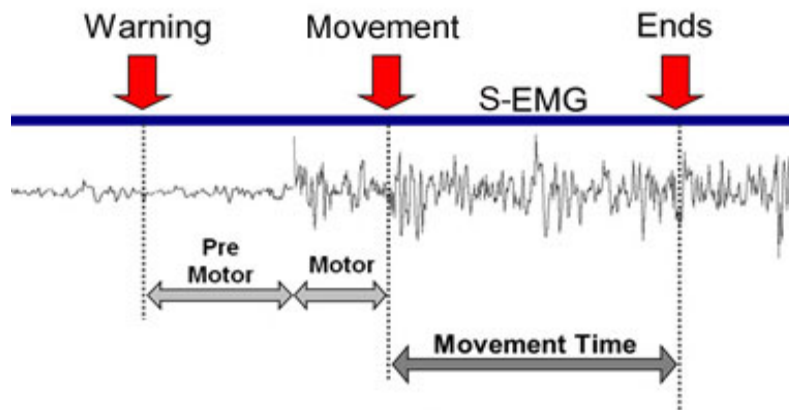

Fig 8. Critical events involved in the reaction time paradigm.

References:

1. Astrand P.O., Rodahl K, Textbook of work physiology, Second Edition. McGraw-Hill, New York, 1977.
2. Basmajian J., DeLuca C.J., Muscles Alive, Fifth Edition, Williams& Wilkins, Baltimore, 1985.
3. Milne-Brown H. S. Stein R. B. , YemmR., Changes in firing rate of human motor units during linearly changing voluntary contraction. J. Physiol 230 (1973) 371-390.
4. Haenneman E., Somjen G.,Carpenter D. O., Excitability and irritability of motoneurons of different sizes. J. Neurophysiology 28 (1965) 599-620.

FUNDAMENTALS TO MAGNETIC RESONANCE (NMR)

AND MAGNETIC RESONANCE SPECTROSCOPY (MRS)

Nuclear Magnetic Resonance (NMR)

A chemical compound consists of molecules, and molecules are composite systems of atoms. An atom consists of a small nucleus and a cloud of electrons. The nucleus is made up from two types of subatomic particles, the protons (p), and the neutrons (n). According to quantum mechanics these subatomic particles are intrinsically spinning. When a number of these particles (p, n) are grouped together to form nucleus, their respective spins will add and the nucleus will have a net nuclear spin. The net nuclear spin is zero for all the nuclei except those with an odd number of protons and an even number of neutrons (and vice versa). These are the nuclei of importance to nuclear magnetic resonance (NMR). As the nuclei spin, their charges circulate and generate a magnetic field. Such magnetic nuclei, which have north and south magnetic poles, have no preferred orientation in space. But if we put them in a uniform static magnetic field, **H** they tend to line up with the field (favorable state). The next thing we do is to change the orientation of the nuclei (perturb the nuclei) in the field (turn them over to make them point the other way). To achieve this (less favorable) state we have to apply energy into the system. This energy can be obtained from the application of a precisely tuned pulsed radio frequency (**RF**) field which is generated from a radio transmitter by changing its frequency. This field is orthogonal to the static field H. When the RF of the transmitter becomes equal to the frequency of the spinning nucleus then we achieve resonance and the RF at which resonance occurs is known as resonance or **Larmor frequency ω**. The equation

$$\omega = -\gamma H$$

is the key equation in NM / MRS, where **H** is the magnetic field strength ,and γ is the **gyromagnetic ratio** (which is associated with each nuclei).

Magnetic Resonance Spectroscopy[1]

MRS has been used by chemists for many years for the analysis of chemical compounds. In medicine[2], MRS is a powerful technique which allows access to the chemistry of the brain or other parts of the human body.

Magnetic **R**esonance **I**maging (**MRI)** images consist of a series of **T1** and **T2** weighted images (T1 and T2 are known as **relaxation times**). They are used to guide the localization of area of interest for MRS studies. That is, from these images, a cube like region (a "voxel" or volume element) is chosen for specific MRS examination.

For example, in the case of: an Alzheimer's disease examination, we might choose to examine an homogeneous area in the occipital grey matter; or for an examination of a tumor, one would obviously choose the site of the tumor.

MRS performs localization through frequency and phase encoding in the presence of a magnetic gradient inside the homogeneous magnetic field of the MRI machine. This allows us to excite the specific region of the body that we are interested in by a **radio frequency** (**RF**) pulse. When one strikes that specific region of the body with a RF pulse, this region begins to resonate based on the chemicals within that region. Then, the resultant resonances are read out using radio frequency detectors. Note that in the brain multiple chemicals resonate at multiple frequencies.

The resultant signal, detected by the detectors, is termed **free-induction decay (FID)**. The FID is a signal generated by the alterations of the local magnetic field, with its amplitude decaying gradually as the magnetization of the region returns to its baseline and is losing its strength.

The FID is a sum of all resonances. Then, by applying Fourier transformation (FFT) on FID, it is possible to determine which chemicals are present (from the **signal's frequency** or **ppm** content) in the region being examined.

The following is a list of the **major neurochemical markers** available at short-echo times (TE: 15 to 35 ms) using standard single-voxel PRESS sequence and includes:

1. NAA (N-Acetyl-Aspartate), a neuronal marker. **NAA peak resonates at about 2.0 ppm**.

2. Cho (Choline), a marker for myelination, indicates axons break up and degradation. **Cho peak resonates at about 3.2 ppm**.

3. Cr (Creatine), an energy marker. **Cr peak resonates at about 3.00 ppm**.

4. Myo (Myoinositol), a marker for gliosis and glial activity. **Myo peak resonates at about 3.5 ppm**.

5. Lac (Lactate), reflects cell death or anaerobic respiration. **Lac** (doublet) **peak resonates at 1.33 ppm** Lactic acid levels get higher when strenuous exercise or other conditions such as heart failure, a severe infection (sepsis), or shock are taken place. Then, blood flow and oxygen throughout the body becomes lower. Very high levels of lactic acid cause a serious, a sometimes life-threatening condition.

6. Glu (Glutamate and glutamine), neurochemical markers. **Glu/Gln peak resonates at about 2.2-2.4 ppm**. Glutamate is a neurotransmitter that is released by nerve cells in the brain. It is responsible for sending signals between nerve cells, and probably very important in the learning and memory processes. If glutamine is too much this may lead to seizures and the death of brain cells, if it is too little, it can cause psychosis, coma and death. Excess Glu, not only over stimulates the nervous system, it is also toxic to the brain and can age/degenerate it too quickly and can cause brain damage after stroke. Glutamine is an important amino acid with many functions in the body. It is a building block of protein and critical part of the immune system. Furthermore, it has a special role in intestinal health. Our body produces this amino acid, and it is also found in many foods.

7. Lip (Lipids), fat in head in pediatrics indicate poor outcome. They are usually associated with necrosis, growth arrest, inflammation, malignancy, apoptosis, and craniopharyngioma which is connected to high amounts of cholesterol in the cyst fluid. **Lip (CH2n) peak resonates at about 1.3 ppm, and Lip (CH3) at 0.9 ppm.**

8. Alanine (doublet) **peak resonates at 1.47 ppm**.

9. GABA (Gamma-aminobutyric acid). **GABA peaks resonate at 2.2-2.4 ppm**.

10. Citrate peak resonates at 2.6 ppm.

These markers are identified by the frequencies (or ppm) at which they occur.

Looking at the list of neurochemical markers just presented, it is surprising that different chemical markers resonate at the same resonance frequencies (ppm). We believe that this is due to (i) the MRI/MRS machine's inability to differentiate them and obtain finer data and (ii) the weakness of FFT to differentiate frequencies that are to close together. More specifically, several frequencies (which correspond to the different chemicals) are close together and fall under an envelope and appear to be as one frequency (but actually there are more). We believe that the new MRI scanners (3 Tesla and above) will resolve the first problem and that Wavelets, can easily resolve the second problem (see the proposed method below).

Note: Head injuries are perhaps the most damaging of all traumas, because of the variability in cause, extent, and effect. They are also, one of the most difficult types of trauma to diagnose. Not too long ago, MRS has emerged as a new and more accurate method in both the diagnosis of

the severity of head trauma, and in the prediction of the outcome of patients, especially in cases where patients are comatose.

At the Huntington Magnetic Resonance Spectroscopy Unit, in Pasadena CA, Dr. B. Ross and his collaborators have reported that in traumatic head injury, many chemical markers have been analyzed and are used for diagnosis. More specifically: They reported that:

- **NAA**, in most cases of head trauma is reduced to some degree. They observed that if this reduction is slight, this indicates optimistic prediction. However, if it is high, this indicates permanent brain damage and severe mental retardation,
- **Cho**: If elevated levels of **Cho** are observed, seen during the breakdown of myelin in axons and in membrane degeneration, associated with diffuse axonal injury, the prognosis is good if Cho is the only abnormal metabolite in the spectrum,
- **Lac**: the presence of **Lac**, is a sign of hypoxia, and in general predicts a very poor outcome for a patient.

Through examination of these metabolites and others, it is possible to obtain a much more quantitative diagnosis of head injury and a much more accurate prediction of future outcome than other clinical standards. Note that in the above studies, the data were phase corrected and the signals from pure water were suppressed by post process.

---

[1] A bit simpler description of the proposed process: When MRS is used the patient is placed inside a homogeneous magnetic field. This field causes the spins of protons to align in a specific direction, designated the longitudinal direction. A short RF pulse transverse to this field is then used to synchronize the precession of these proton spins. When this pulse ends, the spins revert back to their original state, emitting radio signals in the process. The exact signal a proton emits depends on the specific chemical environment of the proton; nuclei in different environments emit radio signals at different frequencies. These signals are combined to form a free induction decay which is then analyzed through FFT. By examining the spectra in frequency space, it is possible to determine the concentrations of certain chemicals and metabolites in a given sample. This makes spectroscopy an extremely powerful diagnostic tool. Furthermore, because the fields used function at radio frequency, magnetic spectroscopy examinations are also extremely safe.

[2] In MRS experiments, it has been observed that high resolution spectral analysis requires relatively low molecular weight compounds (otherwise the spectra become too complex), very pure homogeneous samples, and extremely homogeneous magnetic fields of high strengths (i.e. 3 Tesla or more).

In MRS studies of humans we wish to determine accurately, the chemical composition of a specific region in the human body containing the tumor under study In general, these regions are tissues comprised of very complex molecules, are highly non-

homogeneous, and contain high levels of water and fat as well as small amounts of metabolites (which have been reported to be useful in tumor characterization). Furthermore, the high intensity signals from water and fats severely interfere with the observation of the weak signals from low molecular weight metabolites.

For example, the tissue -water signals is typically four orders of magnitude more intense than that of the metabolites, making it difficult to observe the weak metabolite signals in the present of the intense water signal.

Clearly, the above imposed limitations by the biological systems require strong magnetic fields (i.e. 3 Tesla or more).

FUNDAMENTALS OF WAVELET TRANSFORM (CONTINUOUS AND DESCRETE) AND APPLICATIONS TO MEDICAL SIGNALS

Introduction

Surface electromyography (S-EMG) is a method used to detect the electric potential generated by muscle contraction that is self-generated or caused by external excitation such as the application of a sudden load. S-EMG signals are obtained using a data acquisition system, which acquires and stores the electrical potential responses. These kinds of signals are complex and reflect the physiological properties of a muscle; they are noisy, and non-stationary. Nowadays, despite certain deficiencies of FFT, in most cases, S-EMG signals are analyzed using FFT based methods. During the last decade, in order to resolve these difficulties, another more suitable approach has been introduced. This approach, known as Wavelet transform (WT), is designed to handle this kind of signals. It can help, to de-noise them, to provide high resolution their frequency content, and help us identify the existence of features undetected by FFT methods. Literally, it acts as a mathematical microscope. In this study, we compare several categories of S-EMG signals using Fast Fourier Transform (FT/FFT), Short Time (Window) Fourier Transform (STFT), and Continuous Wavelet Transform (CWT), and the strengths and weaknesses of these methods are discussed. The results show that the FFT and STFT methods have resolution limitations while the CWT approach does not and is particularly useful when time-frequency analysis is important.

1. Mathematical Basis

1.1 From Generalized Fourier series to Wavelet Transforms

Let us consider the following set of the mutually orthonormal vector functions defined in the interval $a \leq t \leq b$

$$\phi_0(t),\ \phi_1(t),\ \phi_2(t),\ \phi_3(t),\ \ldots,\ \phi_n(t),\ \ldots$$

known as "basis functions".

Then, a square integrable function $s(t)$, (a signal of finite energy)

27

$$\int_a^b (s(t))^2 dt < +\infty$$

can be represented by

$$s(t) = c_0\,\phi_0(t) + c_1\phi_1(t) + c_2\phi_2(t) + c_3\phi_3(t) + \ldots + c_n\phi_n(t) + \ldots \tag{1}$$

where
$$c_k = \langle s(t), \phi_k(t)\rangle = \int_a^b s(t)\phi_k(t)\,dt \tag{2}$$

$$k = 0, 1, 2, 3, \ldots, n, \ldots$$

The expression (1) for $s(t)$ is known as generalized Fourier series (GFS) and c_k are known as generalized Fourier coefficients (GFC).

If in the expression (1) for $s(t)$, we use the first n – terms, we obtain an approximation for $s(t)$, represented by $s_n(t)$

$$
\begin{array}{ccc}
\text{Projection of } s_n(t) & \text{Projection of } s_n(t) & \text{Projection of } s_n(t) \\
\text{in 0-direction} & \text{in 1-direction} & \text{in n-direction}
\end{array}
$$

$$s_n(t) \cong \left[\int_a^b s_n(t)\phi_0(t)dt\right]\phi_0(t) + \left[\int_a^b s_n(t)\phi_1(t)dt\right]\phi_1(t) + \ldots + \left[\int_a^b s_n(t)\phi_n(t)dt\right]\phi_n(t) \tag{3}$$

The terms inside the brackets are the known GFC for $s_n(t)$.

In the case that the signal $s_n(t)$ is a non-periodic one, the interval $a \leq t \leq b$ is replaced by the interval $-\infty < t < \infty$, and the basic functions are replaced by trigonometric functions, then the coefficients c_k became the **Fourier Transforms** (FT) of the signal.

1.2 The Problem with Fourier Transform

Although Fourier transform is one of the most important mathematical tools used in the analysis of non-periodic (periodic with infinite period) signals, it has a serious drawback. In transforming $s(t)$ into the frequency domain, time information is lost. That is, when looking at a Fourier transform of a signal, it is impossible to tell when a particular event took place.

If a signal doesn't change over time (stationary signal), this drawback is not very important. However, S-EMG signals contain numerous non-stationary or transitory characteristics such as

frequency bursts, spike discontinuities, abrupt changes, and transients. These characteristics are often the most important part of the signal, and Fourier analysis is not suited to detect them.

In addition, the existence of closely located frequencies can complicate the analysis by eliminating the ability to distinguish one frequency from another requiring the use of carefully designed filters.

1.3 Short Time Fourier Transform (STFT)

Due to the limitation of Fourier Transform on a non-stationary signal, Dennis Gabor (1946) adapted FFTs for the analysis of small sections of the signal at a time (windowing the signal). Gabor named this approach Short-Time Fourier transform (STFT) or Window Fourier Transform (WFT)).

The STFT represents a sort of compromise between time and frequency-based views of a signal, and it provides some information about both (time and frequency) when and at what frequencies a signal event occurs.

In STFT the signal $s(t)$ must have features that are approximately constant in a short time interval, which is the domain of window $w(t)$. If the signal $s(t)$ is multiplied by the window function $w(t)$, a window portion of the original signal $s(t)$ is created, subsequently named $g(t)$. Then, if we translate the window $w(t)$ along the t-axis, we generate the window portions of the original signal $s(t)$: $g(t-t_0)$, $g(t-2t_0)$,… (Fig 1).
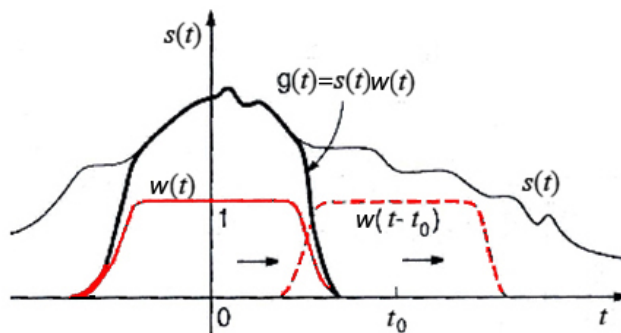

Fig 1. **STFT** windowing process (Ref. [1])

The following is the mathematical description of the STFT process of a non-stationary signal $s(t)$.

29

Continuous STFT: $F(t, \omega) = STFT(t, \omega) = \int_{-\infty}^{\infty} s(\tau)w(\tau - t)e^{-i\omega t}d\tau$ (4)

where $s(t)$ is considered to be stationary within the width of the window $w(t)$, and $\omega = 2\pi f$. Observe that we can only obtain information with limited precision, and the precision is determined by the size of the window $w(t)$. Then, we slide the window along the time axis, while we increase or decrease its width for the particular time t. Then, in each case, we compute the FFT (DFT) of the frame and generate its frequency spectrum.

1.4 The Heisenberg Uncertainty Principle and the Weakness of STFT

The time-frequency resolution is limited by the Heisenberg Uncertainty Principle, which states that for any transform pair

$$w(t) \leftrightarrow W(\omega), \quad \Delta t \cdot \Delta \omega \approx 1$$ (5)

where ω represents frequency.

with $\sigma_t \cdot \sigma_\omega \geq$ constant, at every point $(\sigma_t, \sigma_\omega)$ (Heisenberg Uncertainty Principle)

where: σ_t, σ_ω measure the root mean square (RMS) spread of $w(t)$, and $W(\omega)$ respectively, and are given by

$$\sigma_t = \sqrt{\frac{\int_{-\infty}^{\infty} t^2 |w(t)|^2 dt}{\int_{-\infty}^{\infty} |w(t)|^2 dt}}, \qquad \sigma_\omega = \sqrt{\frac{\int_{-\infty}^{\infty} \omega^2 W|\omega|^2 d\omega}{\int_{-\infty}^{\infty} |W(\omega)|^2 d\omega}}$$ (6)

The product of the resolution in time and resolution in frequency is limited by increasing resolution in frequency, decreasing resolution in time and vice versa. The choice of the window function has a significant effect on the STFT. The window determines how much of the signal will be used in the analysis and controls the frequency resolution of the Fourier spectrum. Such time-frequency resolution is shown below:
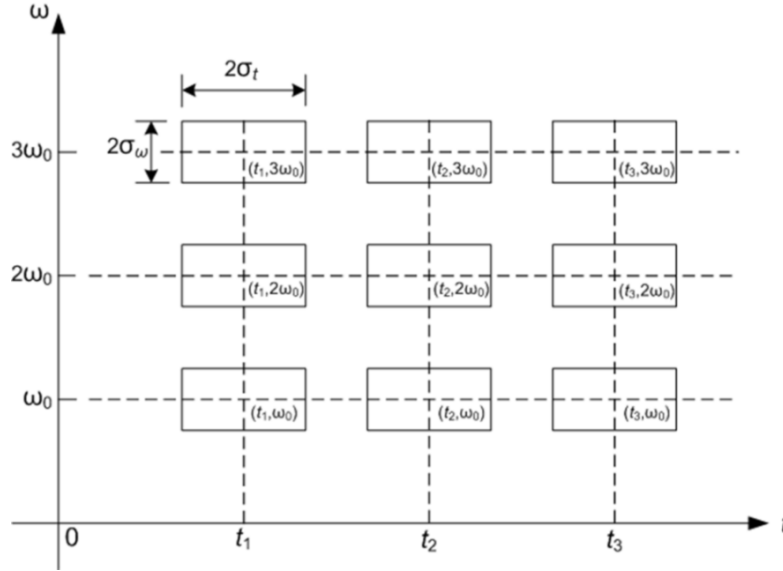
Fig 2. Time-frequency plane showing resolution cells for STFT ($\Delta t = 2\sigma_t$, $\Delta\omega = 2\sigma_\omega$) (Ref.[2])

In addition, if $s(t)$ has a transient component which has a duration smaller than σ_t, it is difficult to locate it with better precision than σ_t. That is since $s(t)$ has key features of varied sizes, it is difficult to find an optimum window function $w(t)$ for the analysis of $s(t)$. Thus, STFT is more suitable for the analysis of signals where all the features appear approximately at the same scale. In the time-frequency diagram shown in Fig 2, for each point (t, ω), we associate a square, centered at that point with side lengths $2\sigma_t$ *and* $2\sigma_\omega$, ($\Delta t \cdot \Delta\omega$, *is defined as* a **resolution cell** (**RC**). In Fig 2, observe that for each point (t, ω), the RC are the same.

In summary, the STFT represents a compromise between time and frequency-based views of a signal. It provides some information about (both) the time and frequency occurrence of an event. However, we can only obtain this information with limited precision, and the precision is determined by the size of the window. Furthermore, the application of STFT is not practical (time consuming). Therefore, we need a (dynamic-scale dependent) window, which has the highest correlation with all the local features of the signal (weakness of STFT).

1.5 The Wavelet Transform (WT)

WT may be thought of as a mathematical microscope which can detect singularities or abrupt changes in the signal as well as fine frequency details. These characteristics of the wavelet

31

transform to make it a more powerful tool than Fourier methods for signal analysis, providing an improved time-frequency resolution.

1.6 Continuous Wavelet Transform (CWT)

CWT is a better alternative to STFT. CWT is designed to overcome the fixed window (resolution problem) of the STFT. In CWT the scale parameter is used. Scale enables us to use a practically infinite number of windows. In CWT, the window is dynamic, covering the low and high frequency information at the same time.

1.7 Mathematical description of CWT

If in the expression (3) for *s(t)*, we replace *the basis functions* $\phi_k(t)$ with $\frac{1}{\sqrt{|a|}} \psi_k \left(\frac{t-b}{a} \right)$

then
$$s(t) = c_0 \frac{1}{\sqrt{|a|}} \psi_0(\frac{t-b}{a}) + c_1 \frac{1}{\sqrt{|a|}} \psi_1(\frac{t-b}{a}) + \ldots + c_n \frac{1}{\sqrt{|a|}} \psi_n(\frac{t-b}{a}) + \ldots \tag{7}$$

where:
$$\frac{1}{\sqrt{|a|}} \psi_0(\frac{t-b}{a}), \frac{1}{\sqrt{|a|}} \psi_1(\frac{t-b}{a}), \frac{1}{\sqrt{|a|}} \psi_2(\frac{t-b}{a}), \ldots , \frac{1}{\sqrt{|a|}} \psi_n(\frac{t-b}{a}) \ldots$$

are mutually orthonormal (2 parameter family) special basis functions defined in $t_1 \leq t \leq t_2$. The coefficients c_n, for n=0, 1, 2, …, can be obtained from expression (2)

$$c_n(a,b) = \frac{1}{\sqrt{|a|}} \int_{t_1}^{t_2} \psi_n(\frac{t-b}{a}) \, s(t) dt \tag{8}$$

If the functions $\psi_k \left(\frac{t-b}{a} \right)$, with $k = 0, 1, 2, 3, \ldots, n, \ldots$ have oscillatory behavior with amplitudes that rapidly decay to zero in both positive and negative directions, then the functions $\psi_k \left(\frac{t-b}{a} \right)$ are called Wavelets.

If in the expression (8) we replace the interval $[t_1, t_2]$ with $(-\infty, \infty)$, then

$$W(a,b) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{\infty} \psi \left(\frac{t-b}{a} \right) s(t) dt \, , \; a \neq 0 \tag{9}$$

This defines the **Continuous Wavelet Transform (CWT)**. The CWT is expressed as the correlation between the signal and the scaled wavelets.

The following is a special mother wavelet known as **Morlet wavelet**

$$\psi_k\left(\frac{t-b}{a}\right) = \pi^{-1/4} e^{i\omega_k\left(\frac{t-b}{a}\right)} e^{-\left(\frac{t-b}{a}\right)^2/2} \quad , k = 0, 1, 2, \dots, n, \dots \quad (10)$$

1.8 Time-frequency relation

For the case of the Morlet wavelet, the following relation connects scale (a) and frequency (ω) (Ref. [6]),

$$a\omega = \frac{\omega_0 + \sqrt{(2+\omega_0^2)}}{2\pi} \quad , \quad \text{where } \omega_0 \geq 5 \; (\omega_0 \text{ is a reference frequency}) \quad (11)$$

1.9 Scalogram

A scalogram is a visualization of a continuous wavelet transform. There are three axes: x representing time, y representing scale, and z representing coefficient value. The z axis is often shown by varying the color or brightness.

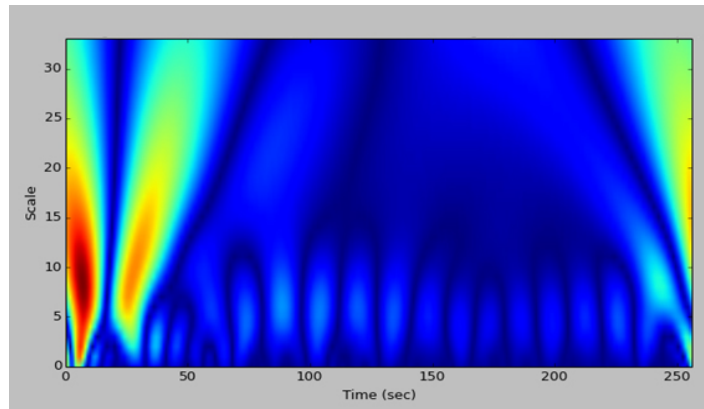A wavelet scalogram is the analogue of a spectrogram for FFT.

Fig 3. S-EMG scalogram.

As stated earlier (see STFT), if a signal $s(t)$ has a transient component which has a duration smaller than σ_t, it is difficult to locate it with better precision than σ_t. In other words, if the signal $s(t)$ has key features of varied sizes, then we cannot find an optimum window function $w(t)$ for the analysis of $s(t)$. However, in the case of CWT, this is possible since for a given time, and

33

time window width, we can have a great number of window choices with various resolution choices in time or scale (frequency). This is clearly shown in Figures 3 and 4.
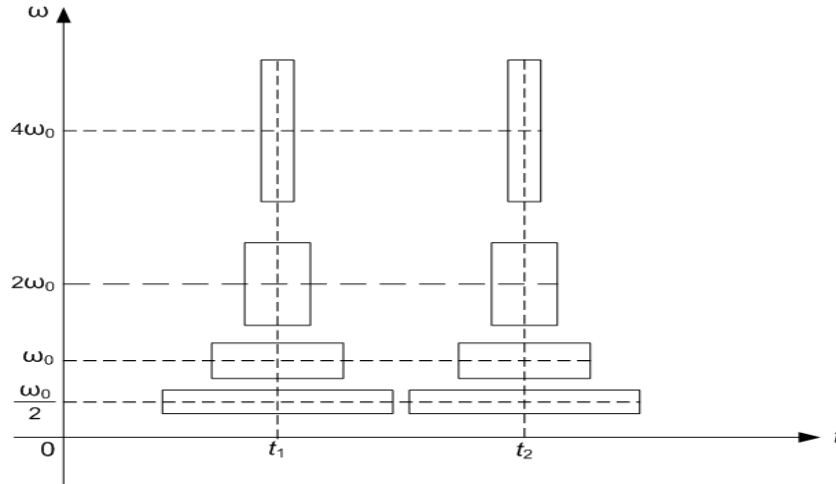

Fig 4. Time-frequency plane showing variable resolution cells for any given time t (Ref. [2]).

1.10 Examples of a typical S-EMG signal and its FFT , STFT, and CWT
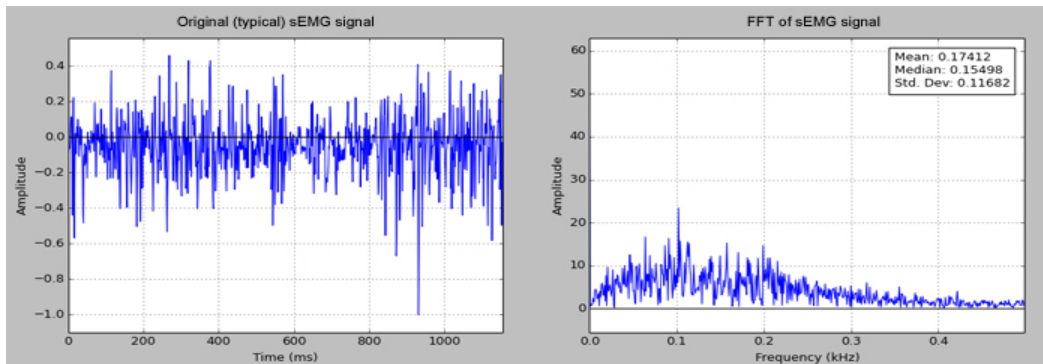
- **FFT of S-EMG**

Fig.5 Fast Fourier Transform of a typical S-EMG signal

When FFT of the entire signal is used, it is impossible to distinguish frequency differences between peeks that are too close to each other. This may cause an inaccurate estimation of the frequency content. Furthermore, in order to understand better, the S-EMG signal, it is necessary to obtain its high-resolution frequency profile. For that reason, traditionally we apply various low pass, high pass, band pass and specially designed filters.

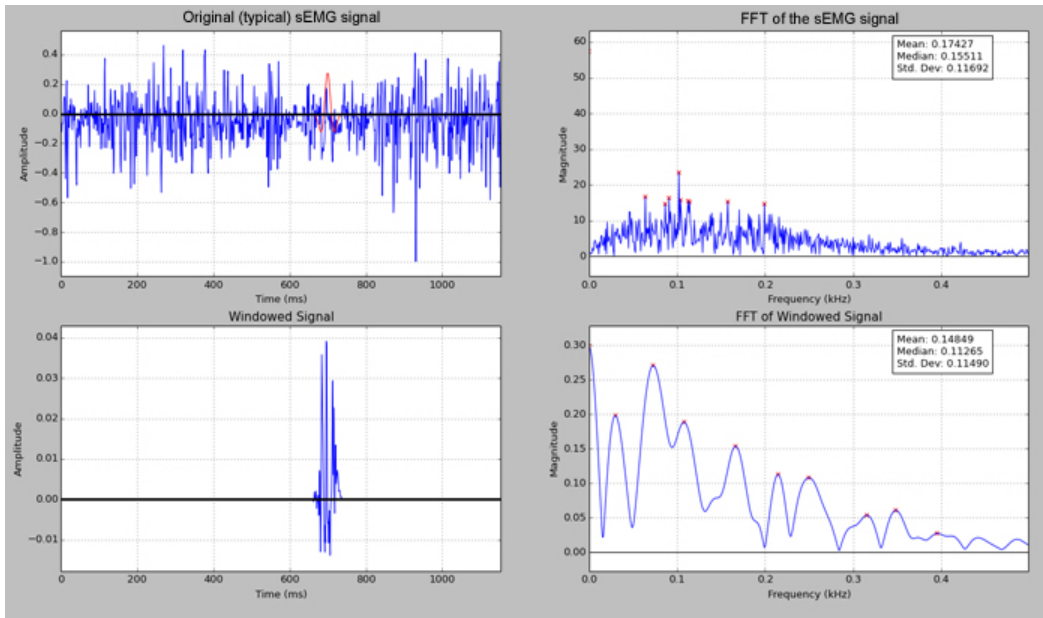- **Example of STFT (WFFT) of S-EMG using Mexican Hat Window**

Fig 6a. Short Time Fourier Transform of a typical S-EMG signal
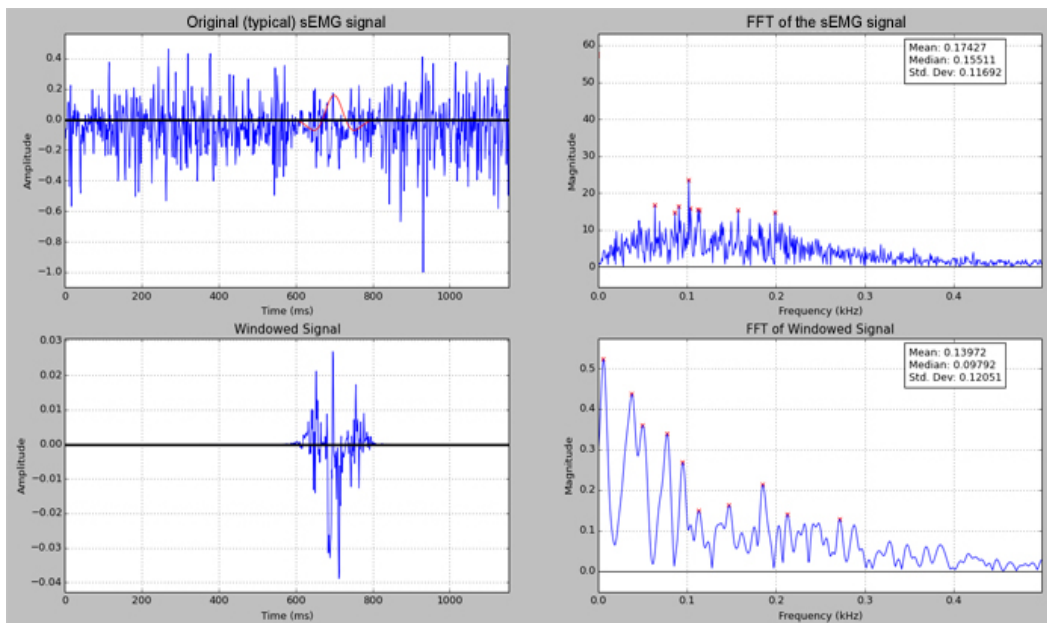using a Mexican Hat window center: 700, width: 100

Fig 6b. Short Time Fourier Transform of a typical S-EMG signal
using a Mexican Hat window center: 700, width: 300

Clearly, the STFT approach helps to resolve the issue of obtaining time and frequency information simultaneously. However, it suffers from constrains imposed by the window width used. Therefore, to resolve this problem a great number of window types and widths are needed, and the optimum window choice is difficult to obtain. More specifically, if a signal $s(t)$ has a transient component which has a duration smaller than σ_t, it is difficult to locate it with precision better than σ_t. In other words, if the signal $s(t)$ has key features of varied sizes, then it is difficult to find an optimum window function $w(t)$ for the analysis of $s(t)$, which is a time-consuming problem. Furthermore, in the STFT approach, it is impossible to distinguish frequency differences between peaks that are too close to each other within the window width. This may cause an inaccurate estimation of the frequency content.

By using the CWT approach, the problem discussed in 2.0 is eliminated. Using CWT (Mexican Hat Window), it is possible that for any given time, we can view the signal with any desired resolution with respect to time and frequency and thus obtain any features that the signal may contain in accordance with the Heisenberg's uncertainty principle. This is clearly shown in Fig.7a, 7b.

- **Examples of the application of CWT on S-EMG (different high-resolution frequency bands)**
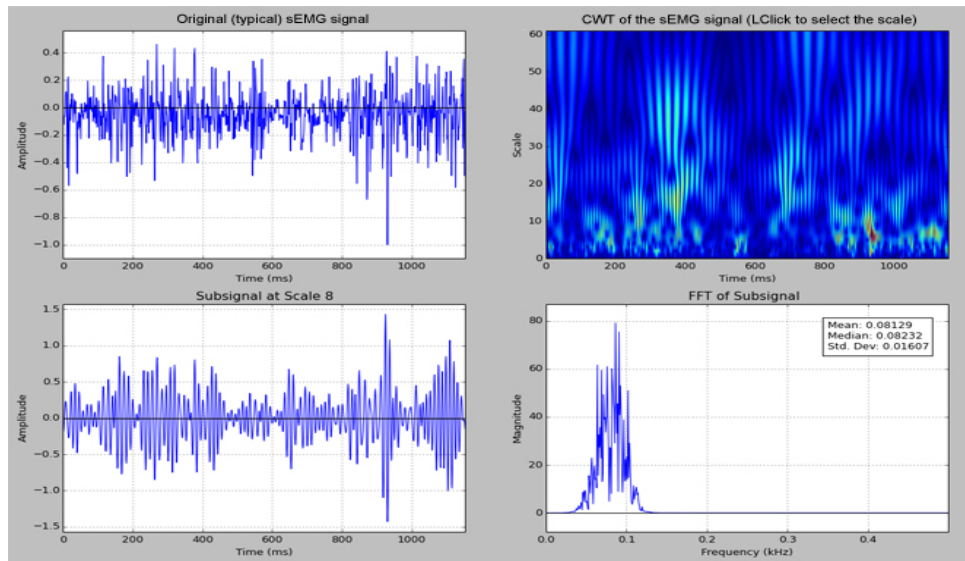

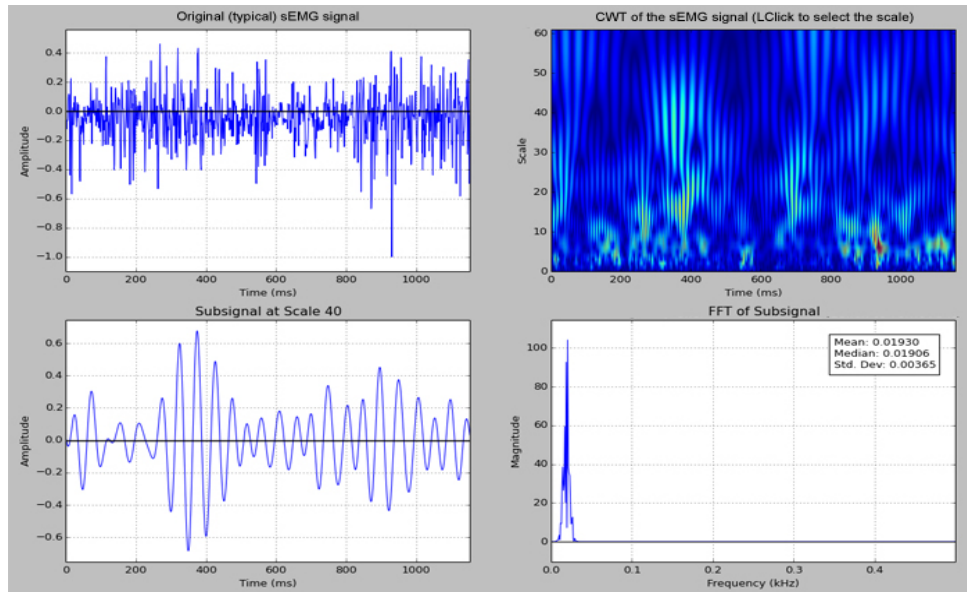Fig.7a CWT of S-EMG, Wavelet Transform

Fig.7b CWT of S-EMG, Wavelet Transform

References

[1] Ingrid Daubechies, "Ten Lectures on Wavelet", Society for Industrial and Applied Mathematics, 1992.

[2] Ali N. Akansu, Richard A. Haddad, "Multiresolution Signal Decomposition" (Transforms, Subbands, Wavelets). Academic Press, Inc. 1992.

2. The Discrete Wavelet Transform, Multiresolution analysis (MRA)

Multiresolution analysis (MRA) Haar wavelet transform (WT) has been applied on a surface electromyography signal (S-EMG). Then, the original signal was decomposed into a predetermined number of sub signals which in turned were subjected to FFT. This resulted into a high-resolution frequency spectrum (frequency bands) for the original signal. The wavelet based spectral analysis was the only way to obtain such a detailed frequency spectrum.

In order to provide this information, the S-EMG signal is decomposed into a predetermined number of sub signals which in turn are subjected to FFT. The resulted frequency spectrum helps us to determine the high-resolution frequency behavior for the S-EMG we are dealing with.

2.1 Theoretical basis

The Haar Decomposition and Reconstruction Wavelet Theory (Ref. 1)

The Haar decomposition and reconstruction transform (averaging and difference algorithm) is demonstrated below using the simple example of 8 data points $s = (s_0, s_1, s_2, s_3, s_4, s_5, s_6, s_7)^T$:
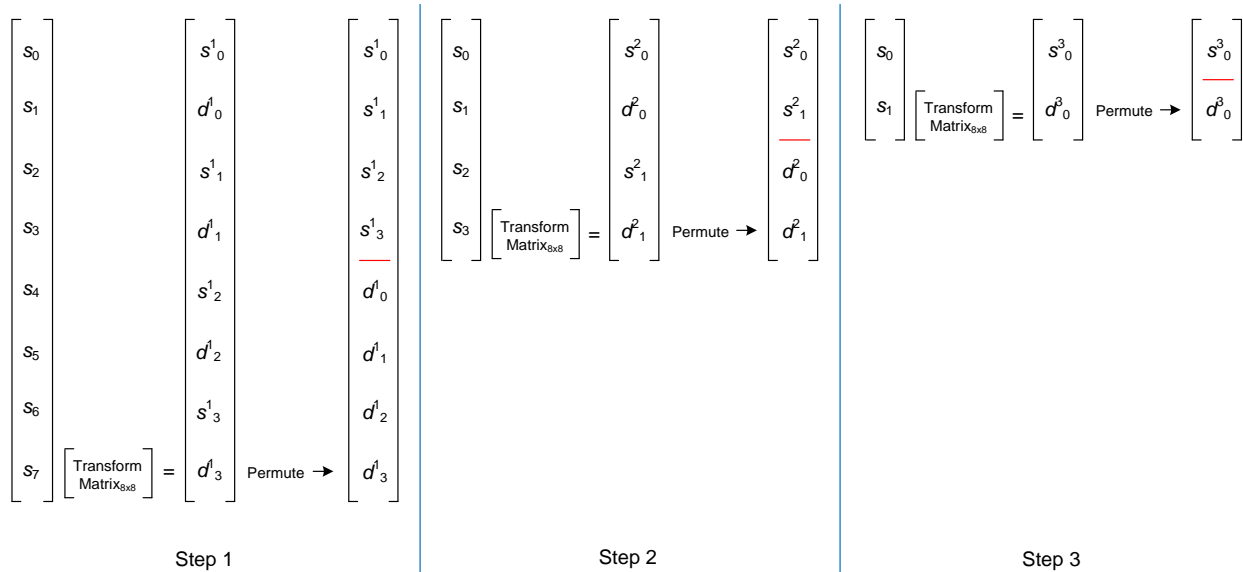

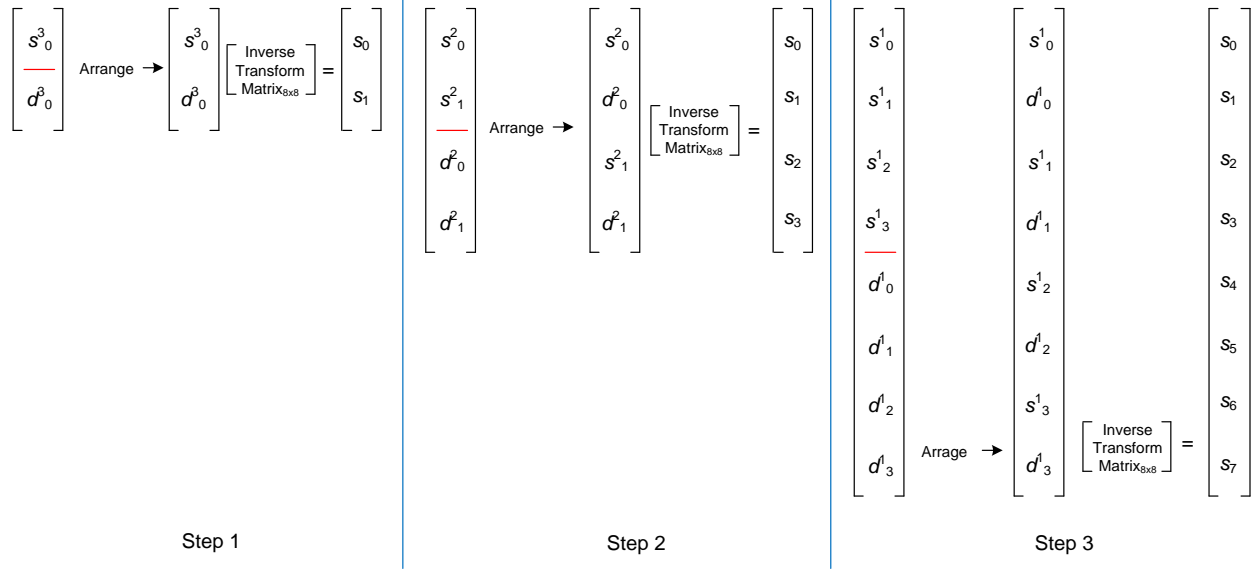Fig. 1 The multiresolution decomposition

Fig. 2 The multiresolution reconstruction

Next, to make things clearer, let us consider the discrete signal $s(t)$, used earlier (2^3 data points) in Hilbert space $L^2([0,1])$. The MRA of $L^2([0, 1])$ consists of a sequence of embedded closed subspaces V_0, V_1, V_2 and V_3 in $L^2([0, 1])$ satisfying certain mathematical conditions (we do not include them here).

The MRA decomposition and reconstruction process can be also depicted from the following diagrams.

where

$$S_L(t) = \sum_{k=0}^{3} c_{2k}\Phi(2^2 t - k), \qquad c_{2k} = \int_0^1 S_L(t)\Phi(2^2 t - k)dt \qquad \text{(decomposition level 1)}$$

$$S_{LL}(t) = \sum_{k=0}^{1} c_{1k}\Phi(2t - k), \qquad c_{1k} = \int_0^1 S_{LL}(t)\Phi(2t - k)dt \qquad \text{(decomposition level 2)}$$

$$S_{LLL}(t) = c_{00}\Phi(t), \qquad c_{00} = \int_0^1 S_{LLL}(t)\Phi(t)dt \qquad \text{(decomposition level 3)}$$

and

$$S_H(t) = \sum_{k=0}^{3} d_{2k}\psi(2^2 t - k), \qquad d_{2k} = \int_0^1 S_L(t)\psi(2^2 t - k)dt \qquad \text{(decomposition level 1)}$$

$$S_{LH}(t) = \sum_{k=0}^{1} d_{1k}\psi(2t - k), \qquad d_{1k} = \int_0^1 S_{LH}(t)\psi(2t - k)dt \qquad \text{(decomposition level 2)}$$

$$S_{LHH}(t) = d_{00}\psi(t), \qquad d_{00} = \int_0^1 S_{LLH}(t)\psi(t)dt \qquad \text{(decomposition level 3)}$$
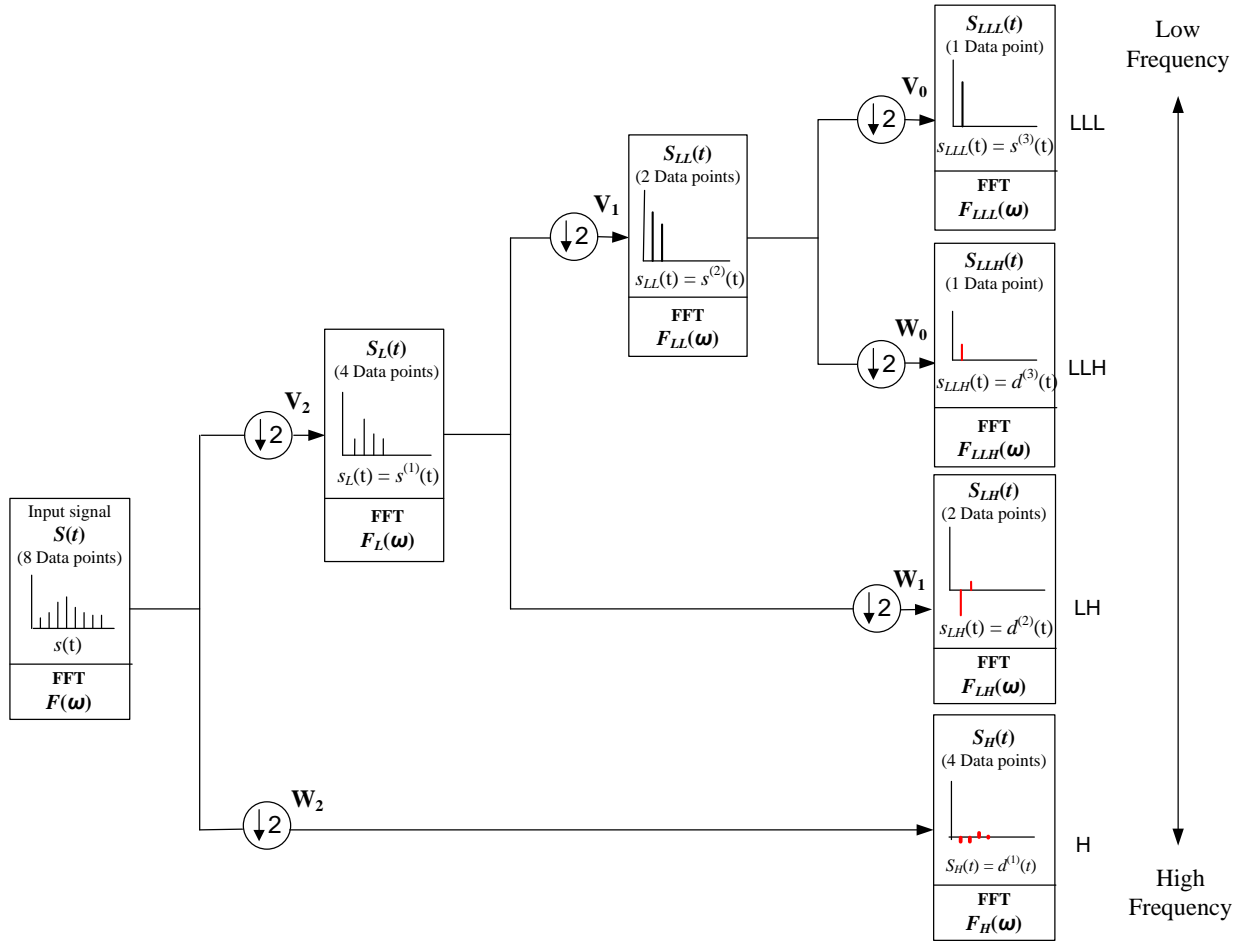
Fig.3 Decomposition Process Chart for 8 data points

The MRA reconstruction process is the inverse process of the decomposition. It recovers the original signal $S(t)$ as follows.

$$S(t) = S_{LLL}(t) + S_{LLH}(t) + S_{LH}(t) + S_H(t)$$

where $S_{LLL}(t) \in V_0$, while $S_{LLH}(t) \in W_0$, $S_{LH}(t) \in W_1$, and $S_H(t) \in W_2$.

$$S(t) = c_{00}\Phi(t) + d_{00}\psi(t) + \sum_{k=0}^{1} d_k\psi(2t - k) + \sum_{k=0}^{3} d_{2k}\psi(2^2t - k)$$

where $S(t)$ will be $S_{LLH}(t)$, $S_{LH}(t)$, and $S_H(t)$ depending on the level of reconstruction.
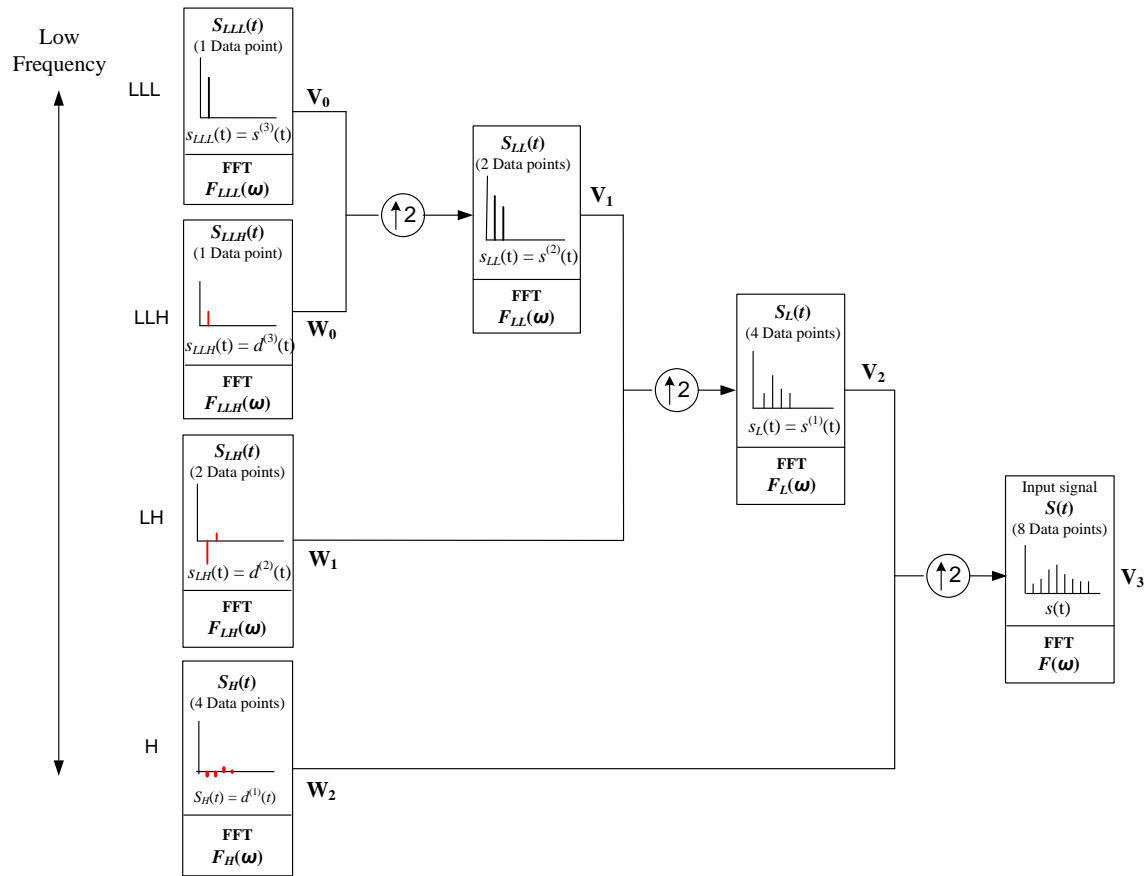
Fig 4. Reconstruction Process Chart for 8 data points

The **Haar transform** is the simplest of the wavelet transforms. This **transform** cross-multiplies a function against the **Haar** wavelet with various shifts and stretches, like the Fourier **transform** cross-multiplies a function against a sine wave with two phases and many stretches. From the analysis of the experimental data via the Haar Transform to analysis the S-EMG signals, we observe that the mean frequency values for all decomposition levels (L7 and L8) is between 0.07 - 0.21, and that the two highest levels contribute the highest frequencies. Clearly, one can observe from Fig.3 that the FFT of the original signal does not provide a high-resolution spectrum.

2.2 The Daubechies (D4) Decomposition and Reconstruction Wavelet Theory (Ref. 1) and Mallat's Algorithm (Ref.2)

The discrete wavelet transform provides a method for the analysis of S-EMG signals. It allows specific features of a signal to be localized in time by decomposing the signal into a family of basis functions of finite length, called wavelets. A particular property of the method is its ability to identify and isolate fine structures of a signal. This may be the small perturbations in an otherwise smoothly varying signal which are difficult or impossible to detect by other means. That property makes the discrete wavelet transform valuable for the analysis of S-EMG signals.

Discrete Wavelet Transform (DWT), the Pyramidal Algorithm (Ref. 3).

A signal $s(t)$, defined in the interval $0 \leq t < 1$, is represented by the discrete sequence

$$s(t) = \left(s_0(t), \ s_1(t), \ s_2(t), \dots, \ s_{n-1}(t_{n-1}) \right)^T \tag{1}$$

The DWT Pyramidal Algorithm (PA) transforms the sequence defined in (1) into a new sequence

$$a(t) = \left(a_0(t), a_1(t), a_2(t), \dots, a_{n-1}(t_{n-1}) \right)^T \tag{2}$$

$$s(t) = L_3^T L_2^T L_1^T a_0(t) + \ L_3^T L_2^T H_1^T a_1(t) + \ L_3^T H_2^T \begin{bmatrix} a_2(t) \\ a_3(t) \end{bmatrix} + \ H_3^T \begin{bmatrix} a_4(t) \\ a_5(t) \\ a_6(t) \\ a_7(t) \end{bmatrix} \tag{3}$$

where L signifies low-pass filtering and H signifies high-pass filtering. The transformation matrices L_i and H_i, $i= 1, 2, 3$ must be orthogonal, and the coefficients $a_i(t)$ are given by

$$a_0(t) = \frac{1}{2} L_1 \frac{1}{2} L_2 \frac{1}{2} L_3 \ s(t); \qquad a_1(t) = \frac{1}{2} H_1 \frac{1}{2} L_2 \frac{1}{2} L_3 \ s(t)$$

$$\tag{4}$$

$$\begin{bmatrix} a_2(t) \\ a_3(t) \end{bmatrix} = \frac{1}{2} H_2 \frac{1}{2} L_3 \ s(t); \qquad \begin{bmatrix} a_4(t) \\ a_5(t) \\ a_6(t) \\ a_7(t) \end{bmatrix} = \frac{1}{2} H_3 \ s(t)$$

For the case, Daubechies' D4 wavelet (Ref. 1), the transformation matrices are:

$$L_1 = [c_0 + c_2 \quad c_1 + c_3];$$

$$L_2 = \begin{bmatrix} c_0 & c_1 & c_2 & c_3 \\ c_2 & c_3 & c_0 & c_1 \end{bmatrix};$$

$$L_3 = \begin{bmatrix} c_0 & c_1 & c_2 & c_3 & & & \\ & & c_0 & c_1 & c_2 & c_3 & \\ & & & & c_0 & c_1 & c_2 & c_3 \\ c_2 & c_3 & & & & & c_0 & c_1 \end{bmatrix}$$

$$H_1 = [-c_3 - c_1 \quad -c_2 + c_0];$$

$$H_2 = \begin{bmatrix} -c_3 & c_2 & -c_1 & c_0 \\ -c_1 & c_0 & -c_3 & c_2 \end{bmatrix};$$

$$H_3 = \begin{bmatrix} -c_3 & c_2 & -c_1 & c_0 & & & \\ & & -c_3 & c_2 & -c_1 & c_0 & \\ & & & & -c_3 & c_2 & -c_1 & c_0 \\ -c_1 & c_0 & & & & & -c_3 & c_2 \end{bmatrix} \qquad (5)$$

Where
$$c_0 = (1 + \sqrt{3})/4, \quad c_1 = (3 + \sqrt{3})/4, \qquad (6)$$
$$c_2 = (3 - \sqrt{3})/4, \quad c_3 = (1 - \sqrt{3})/4$$

References

1. *Ten Lectures on Wavelets,* I. Daubechies, SIAM, 61, 1992.

2. *A wavelet Tour of Signal Processing,* S. Mallat, Academic Press, 1998.

3. *Some Properties of discrete wavelet maps,* D. E. Newland, Probabilistic

   Engineering Mechanics, 9 (1994), 59-69.

4. *Analysis of S-EMG Signals using Multiresolution Analysis (MRA) Haar

   Transform,* Billy Chung M.S., N. D. Panagiotacopulos, D. H.K. Chow (Forthcoming publication)

5. *Wavelets and Filter Banks* by Gilbert Strang and Truong Nguyen, Wellesley-Cambridge Press,
   1997.

6. *Haar wavelets based technique in evolution problems* by Cattani, Proc. Estonian Acad. Sci.
Phys. Math., 2004, 53, 45–65.

7. *Haar wavelet method for solving lumped and distributed parameter systems* by Chen, C. F.
and Hsiao, C.-H., IEE Proc. Control Theory Appl., 1997.