Signal processing of the surface electromyogram to gain insight into neuromuscular physiology

BY J. STEFAN KARLSSON1,*, KARIN ROELEVELD2, CHRISTER GRÖNLUND1, ANDREAS HOLTERMANN2 AND NILS ÖSTLUND1

1Department of Biomedical Engineering & Informatics, University Hospital, 901 85 Umeå, Sweden
2Human Movement Sciences Programme, Norwegian University of Science and Technology, Dragvoll Idrettscenter, 7491 Trondheim, Norway

A surface electromyogram (sEMG) contains information about physiological and morphological characteristics of the active muscle and its neural strategies. Because the electrodes are situated on the skin above the muscle, the sEMG is an easily obtainable source of information. However, different combinations of physiological and morphological characteristics can lead to similar sEMG signals and sEMG recordings contain noise and other artefacts. Therefore, many sEMG signal processing methods have been developed and applied to allow insight into neuromuscular physiology. This paper gives an overview of important advances in the development and applications of sEMG signal processing methods, including spectral estimation, higher order statistics and spatio-temporal processing. These methods provide information about muscle activation dynamics and muscle fatigue, as well as characteristics and control of single motor units (conduction velocity, firing rate, amplitude distribution and synchronization).

Keywords: electromyography; multi-channel; spatio-temporal processing; neuromuscular physiology; muscle function; muscle control

1. Introduction

Surface electromyography is a commonly applied technique to evaluate physiological and morphological characteristics of muscles and their neural strategies. It involves registration of electrical potentials from active muscle fibres by electrodes placed on the skin above the muscle. More precisely, the recorded signal or the surface electromyogram (sEMG) consists of the weighted sum of the electrical contributions of active motor units (MUs). The MU is the smallest functional unit of movement, and consists of a motor neuron and the group of muscle fibres it innervates. The electrical representation of active MUs, and therefore the sEMG, contains information about the characteristics and physiology of the active MUs (e.g. MU size, activation and firing pattern, and muscle fibre type, size and state). In addition, the sEMG is related to force production and can be used to gain insight into which situations MUs are

* Author for correspondence (stefan.karlsson@vll.se).

One contribution of 13 to a Theme Issue ‘Signal processing in vital rhythms and signs’.
activated, thus enlightening the neural strategy. Furthermore, since detection of this signal involves electrodes placed on the skin above the muscle, it barely interferes with normal muscle activity and body movements. However, this large amount of information in the signal also causes multiple solutions to the inverse problem of determining single neuromuscular properties. In addition, the recorded signal includes noise, interference and other artefacts. The sEMG depends on the electrodes, the tissue between the electrodes and the muscle, and the acquisition equipment. To allow insight into neuromuscular physiology from the sEMG, three indispensable approaches are used: standardization of data collection (Hermens et al. 1999); investigation of the relationship between the signal and physiology (e.g. Roeleveld & Stegeman 2002); and development and application of signal processing methods.

This paper gives an overview of important recent advances of developed and applied signal processing methods in surface electromyography to gain insight into neuromuscular physiology. For a more extensive overview of methods applied in surface electromyography, the reader is referred to Basmajian & De Luca (1985) and Merletti & Parker (2004). After an introduction about the origin of the sEMG (§2), a presentation of some pre-processing methods (§3), signal processing methods for the estimation of muscle activity (especially dynamics of activation, §4), the sEMG spectrum and its descriptors (§5), muscle fibre conduction velocity (CV; §6), single MU activity (decomposition; §7), MU synchronization (§8) and localization of the innervation zone (IZ; §9) are presented. In each section, first, the physiological problem or goal is defined, then the classical solutions are given and then detailed examples of recent methods, mostly developed by our group, are presented. Where possible, first, signal processing methods for traditional single-channel sEMG signals are presented followed by methods for multi-channel or high-density sEMG signals. In multi-channel surface electromyography, multiple electrodes are placed over the skin of a single muscle, while high-density surface electromyography is referred to as a specific case of multi-channel surface electromyography in which relatively small electrodes with small inter-electrode distances are applied.

2. The surface electromyogram

MU action potentials (MUAPs) are the most fundamental elements of the sEMG. The MUAP is the summation of single-fibre action potentials (SFAPs) propagating over the muscle fibres belonging to the MU. The source of the SFAP is the fluctuation in Na\(^+\) and K\(^+\) around the cell membrane, which generates an intracellular action potential that then causes a transmembrane current. Owing to volume conduction, an extracellular potential field can be recorded by extracellular electrodes. As a result of motor neuron firing, an action potential is generated at the motor endplate where the nerve endings connect to the muscle fibre (typically around the middle of the fibre). The action potential propagates with a specific CV towards both ends of the fibre (figure 1). The motor endplates of all muscle fibres within a muscle are typically located within a small band called the IZ.

The SFAPs from each muscle fibre of the same MU fire almost simultaneously. The shape of one individual MUAP detected on the skin is therefore, to some extent, affected by the spatial arrangement of the motor endplates, but also by
the muscle fibre thickness, the chemical characteristics (fibre type), the state of the fibres (fatigue), the depth of the fibres within the muscle and their distances from the detection electrode. Each MUAP contributes to the amplitude and frequency contents of the sEMG (interference pattern sEMG). Since large MUs cause high-amplitude MUAPs, and since tissue acts as a low-pass filter, large MUs and MUs close to the electrode give a higher contribution to the amplitude and frequency contents from a sEMG. A motor neuron usually fires with a certain rate resulting in MUAP trains (figure 1). The firings are not strictly periodic and the firing intervals can be described by a Gaussian distribution. At low contraction levels, the average firing rate can be as low as a few pulses per second (pps), and the rate typically increases with activation level up to approximately 30 pps. With increasing activation level, the number of active MUs also increases. In slow varying, non-fatiguing ramp contractions, such MU recruitment takes place through the size principle: MUs with small neurons are recruited first, after which MUs with gradually larger neurons are recruited. MUs with small neurons generally compose small, slow, weak and relatively low-fatigable fibres. Under most low and moderate contractions, the different MUs fire almost independently from each other. However, with increasing contraction level and, as a result of fatigue, training or some diseases, the firings can become more temporally related, called MU synchronization.

Traditionally, a sEMG is recorded as the difference between two electrodes (bipolar or single differential electrode configuration) placed on the skin above the active muscle, along the fibre orientation, preferably away from the IZ and tendon region. Typically, electrodes have a diameter of approximately 10 mm and an inter-electrode distance of approximately 20 mm. The difference between the two signals is amplified, eliminating the common signals from electrical devices (mainly power mains) and more distant muscles. This gives information...
from a large part of the muscle. Application of multiple electrodes, called multi-channel surface electromyography, using two-dimensional high-density electrode grid systems with small electrodes (approx. 1.5 mm) and small inter-electrode distances (approx. 5 mm), allows a more detailed investigation of the active muscle. In particular, by applying spatial filters to the detection system, the activity of different regions of a muscle can be spatially separated (spatial selectivity) and individual MUAPs can be investigated. In addition, cross-talk, i.e. activity from distant bioelectrical signals, can be reduced. In principle, the spatial filters can be obtained by using different recording configurations. However, to be more flexible, often monopolar configurations (electrodes above the muscle of interest with respect to a common reference placed nearby, above non-active tissue) are used for data collection, and then spatial filtering is carried out after digitalization. See §3b for more information on this topic.

The individual MUAPs tend to have a consistent shape in healthy, non-fatigued persons and can be regarded as quasi-deterministic. However, as the number of independently firing MUs increases, the sEMG is well described as a band-limited Gaussian-distributed stochastic process with zero mean (due to the central limit theorem). The amplitude of the recorded sEMG is typically in the range of a few millivolts (peak to peak), and the frequency spectrum ranges from approximately 10 to 500 Hz, with the dominant energy in the 20–150 Hz range. The sEMG spectrum is influenced by two major factors (recording factors excluded): (i) in the low-frequency range (5–40 Hz), the MU firing rate and (ii) in the high-frequency range (above 40 Hz), the form of the individual MUAPs (Basmajian & De Luca 1985). The sampling frequency is often set at 1 or 2 kHz. In order to avoid aliasing of non-sEMG signals, according to the Nyquist theorem, the analogue signals are low-pass filtered at approximately 500 Hz. Before digitizing, the signal needs signal conditioning, i.e. amplification and filtering. For more detailed information, see Merletti & Parker (2004).

3. Pre-processing methods

The recorded sEMG is influenced by technical factors (acquisition equipment noise, electrode–skin contact properties, etc.) in addition to physiological factors. Therefore, prior to physiological interpretation or further processing, it is necessary to assess whether or not the signal is of sufficient quality. Pre-processing in the form of filtering is usually carried out to reduce the influence of the technical factors. These two important aspects are presented in §3a,b.

(a) Signal quality assessment

Classically, sEMG quality is estimated by the electrode–skin impedance, which is measured prior to signal acquisition. This gives information on the condition of the contact between the electrodes and the skin. However, contact and electrode–skin impedance may change during acquisition. Therefore, it is desirable to base the signal quality assessment on the acquired sEMG signals directly and on short periods of time.

While the signal quality could be assessed by visual inspection, this would be very laborious in long-term measurements of signals and, especially, in multi-channel recorded signals. There are only a few methods in the literature for
automatic detection of low quality of the sEMG. A general approach is to detect and reject outliers of observations (time samples) within a signal. For a single-channel sEMG, Falou et al. (2005) presented a method based on a regression technique, where the normalized signal spectrum is compared with an exponential model during a signal segment.

For multi-channel sEMG recordings, Grönlund et al. (2005a) described an online spatio-temporal approach. The basis for the method is that monopolar signals of a multi-channel recording are expected to lie within a certain range due to the low-pass filtering effect of volume conduction. Therefore, outliers probably represent a signal of non-physiological origin. First, the signal amplitude recorded from each electrode is calculated by its standard deviation values in two time windows of different lengths throughout the recording (for the detection of both intermittent and long-term noise), \(Y_i[n] = \sigma(x_i[m], m \in \{n \ldots n + T_1\})\) \(\sigma(x_i[m], m \in \{n \ldots n + T_2\})\). Then, the variables are de-correlated and normalized, \(\tilde{Z}_i \leftarrow Y_i\). Finally, a signal recorded at channel, \(i\), is classified as noise, or a non-sEMG, if the ‘outlier probability’ \(p_i > 0.05\),

\[
p_i = N^{-1} \sum_n \{||\tilde{Z}_i[n]|| > 1\}. \tag{3.1}
\]

(b) Filtering

The general objective of filtering, whether it is in the time or spatial domain, is to improve the signal-to-noise ratio (SNR) by reducing movement artefacts, high-frequency noise and activity from irrelevant tissue (mostly distant muscles or MUs). This is most commonly achieved by the application of a band-pass filter in the time domain and a high-pass spatial filter (differential filter).

To preserve the physiological signal and reduce movement artefacts, a band-pass filter with low and high cut-off frequencies around, respectively, 10–30 and 300–500 Hz (and sometimes 50 or 60 Hz notch filters) is usually applied to the sEMG. For specific aims, the filtering can be substantially different, e.g. an extreme high-pass filtering in order to improve the force estimation from the sEMG (Potvin & Brown 2004). By contrast, Dolan et al. (1995) used only the low-frequency part of the spectrum to track muscle fatigue.

After data collection, traditional single-channel surface electromyography is restricted to the application of temporal filters. By contrast, high-pass spatial filters can be applied on multi-channel sEMG signals and, especially, on high-density sEMG signals when recorded with a monopolar configuration.

A spatial filtration is a linear combination of signals from different electrodes. If the recording system is larger than the filter kernel, the filtering is performed as a spatial convolution. Thus, the filters can be defined as matrices that consist of their filter coefficients.

The bipolar configuration, also called longitudinal single differential filter, is the most common spatial filter in surface electromyography. However, the spatial selectivity of the bipolar sEMG is often too low to allow single MUAP investigation (Rau & Disselhorst-Klug 1997). Therefore, different types of high-pass spatial filters are used to limit the number of contributing MUAPs. One filter, which is often used for this purpose, is a two-dimensional differentiating spatial filter, often of size 3×3, called the Laplace filter (sometimes called a

Phil. Trans. R. Soc. A (2009)
normal double-differentiating filter). Of course, there are other filters that are
more or less common, and almost all of them are defined as fixed spatial filters,
I.e. the same coefficients are used for all spatial locations (spatial convolution;

The choice of the specific settings of these filters should depend on the
characteristics that are to be extracted in relation to the total signal. Because the
sEMG is also dependent on non-physiological factors and changes with
contraction level and duration, the optimal filter settings are expected to differ
between persons, recordings, contraction level and duration. Therefore, adaptive
filters could be a good alternative to the traditional fixed filters when specific
characteristics are to be extracted.

Filters can be used to extract single MUAP trains. In that situation, it would
be desirable to use a criterion that gives clearly visible peaks. One criterion,
which is sensitive for outliers, is the kurtosis of the signal. Therefore, this
criterion has been used in an adaptive filter for enhancing MUAP peaks (Östlund
et al. 2006), where the filtering is performed with the coefficient that maximizes
the kurtosis of the output. Because the spatial filter is a linear combination, this
is comparable to projection pursuit.

In surface electromyography, the filtering in the time domain is often
separated from the filtering in the spatial domain. However, if the spatial filtering
and time filtering are rewritten as a linear combination of time-delayed versions
of the input signals, the spatio-temporal filtering can adaptively be determined
with the same approach as for the above-mentioned adaptive spatial filter. That
is, the combination of the convolution in the time domain and the linear
combination in the spatial domain can be written as one linear combination,

\[ y[n] = \sum_{i} a_i (h_i \ast x_i) = \sum_{i} a_i \sum_{m} h_i[m] x_i[n-m] = \sum_{i} \sum_{m} w_i[m] x_i[n-m], \]

where \( x_i \) is the \( i \)th input signal, with the time domain filter kernel \( h_i \) and the
spatial filter coefficient \( a_i \).

The filter coefficients in the spatio-temporal filter can be determined by finding
the linear combination of time-delayed input signals that, for example, maximizes
the kurtosis of the output \( y \). Since the kurtosis is a fourth-order cumulant, it
is possible to achieve this by diagonalizing fourth-order cumulants, e.g. with
the joint approximate diagonalization of eigenmatrices (JADE) algorithm. This
can be compared with principal component analysis that uses second-order
cumulants. When used for locating single MUAP trains, an adaptive spatio-
temporal filter had (for simulated signals) a sensitivity of 98.7 per cent and a
positive predictivity of 86.8 per cent, while the Laplace filter had a sensitivity of
64.4 per cent and a positive predictivity of 70.2 per cent (Östlund et al. 2006).

4. Muscle activation dynamics

The amplitude of the sEMG is often used to assess muscle activation and force
level. For this purpose, the sEMG amplitude of single-channel recordings is
traditionally quantified using the root-mean-square (RMS) and the average
rectified value estimators. However, the level of activation is seldom constant in

Phil. Trans. R. Soc. A (2009)
time and homogeneously distributed over the muscle, even at constant force contractions. These spatio-temporal changes in activation indirectly contain information about MU recruitment (Scholle et al. 1992; Holtermann et al. 2005) and may also provide information of plausible motor control strategies to prevent fatigue during prolonged contractions, i.e. intramuscular activity redistribution and differential activation between neuromuscular compartments (Farina et al. 2008; Holtermann et al. 2008a). In addition, these phenomena are gaining popularity in the field of ergonomics to gain insight into mechanisms associated with muscle pain. Therefore, different methods have been developed to investigate and quantify these muscle activation dynamics.

(a) Single-channel analysis

While linear changes in amplitude within a contraction can be analysed by straightforward regression analyses or percentage change calculations, properties of the dynamics of the amplitude distribution of a single signal can be analysed using different methods. Exposure variation analysis is based on constructing a two-dimensional histogram, which gives joint information on the frequency (percentage of total time) of different amplitude levels at different time window lengths. A similar approach is to analyse so-called gaps, a method in which very short incidents of muscle relaxation are analysed.

(b) Multi-channel analysis

The changes in spatial distribution can be described by two-dimensional profiles of the amplitude distribution from two-dimensional multi-channel sEMG recordings.

Some methods using multi-channel techniques have been proposed (Holtermann et al. 2005; Holtermann & Roeleveld 2006; Farina et al. 2008). Farina et al. (2008) used the centre of gravity of two-dimensional RMS maps to quantify global changes in the spatial distribution of muscle activity. Holtermann et al. (2005) used correlation between successive two-dimensional RMS maps, 

\[ R(X_+[m_1], X_+[m_2]) = \frac{\text{cov}(X_+[m_1], X_+[m_2])}{\sqrt{\text{var}(X_+[m_1])\text{var}(X_+[m_2])}} \]  

(4.1)

where \( X_+[m] \) represents the RMS values of all channels at time segment (epoch) \( m \). This provides information about the degree of change in spatial distribution, where a low correlation indicates a large change in spatial distribution, and a high correlation indicates a small change in spatial distribution. By correlating the RMS values from time samples with different levels of force or fatigue, recruitment and/or firing rate modulation and CV changes of MU populations during different tasks and conditions can be obtained (Holtermann & Roeleveld 2006).

Changes in the intramuscular activity can also be used to study so-called differential activations between different neuromuscular compartments. By detrending and normalizing the signals obtained from electrodes (placed transverse to the fibre orientation) with large pickup areas above two different regions of a muscle, so-called differential activations can be analysed (Holtermann et al. 2008a). Differential activations are detected as amplitude peaks above a threshold in the differential signal obtained from those signals.
5. Muscle fatigue (single-channel spectral analysis)

Muscle fatigue has been extensively studied with surface electromyography (Lindström & Magnusson 1977). During a sustained, high-effort muscle contraction, muscle fibres fatigue, causing a decreased CV and a reduced force production per muscle fibre. To maintain the desired force, more MUs have to be recruited. In addition, the firing rate and MU synchronization can be altered. Direct information about these parameters can be obtained by multi-channel recordings (see §§6–8).

All these parameters are more or less reflected in the frequency content of the single-channel sEMG and can therefore be investigated using spectral analysis. During muscle fatigue, the total power of the sEMG (related to the RMS) increases and the power spectral density (PSD) function of the sEMG shifts gradually towards lower frequencies. These changes are mainly caused by recruitment of additional MUs and a decreased CV, respectively.

Because the sEMG frequency content is affected by all these different physiological factors, in addition to non-physiological ones, spectral changes or differences under standardized conditions and exercise protocols are mostly investigated. For example, spectral analysis has been used to estimate changes and differences in CV (Lindström & Magnusson 1977; Basmajian & De Luca 1985), muscle fibre-type proportion (Gerdle & Fugl Meyer 1992; Kupa et al. 1995) and firing rate. Through the estimation of changes in these parameters, spectral analysis has been used for diagnostic classification (Inbar & Noujaim 1984) and to gain insight into MU recruitment strategies (Solomonow et al. 1990) and force production (Bigland-Ritchie 1981; Basmajian & De Luca 1985).

Two essential methodological steps must be performed to use spectral analysis: the estimation of the spectrum itself and the choice of the spectral indicator. The state of the art of these two aspects in single-channel bipolar surface electromyography is given in §5a,b. Of course, these techniques can be applied to multi-channel surface electromyography and spatial dynamics of the sEMG spectrum can be included in the analysis.

(a) Spectral estimation

Several different methods are used for spectral estimation of the sEMG. The most widely used method for PSD estimation of the sEMG is the Fourier transform (FT), partly owing to the extremely computationally effective fast FT (FFT) algorithm. However, the application of the FFT requires a stationary signal, while the characteristics of the sEMG change with contraction level and time (as a consequence of muscle fatigue), i.e. the signal is not stationary. For isometric contractions, this is often solved by dividing the long-term sEMG into a sequence of non-overlapping epochs and estimating a PSD for each. This is because, under those conditions, the sEMG is assumed to be wide-sense stationary for epochs lasting 0.5–2 s, depending on the contraction level and the investigated muscle. Similarly, using isokinetic dynamometers, highly standardized repetitive dynamic contractions can be studied. By determining sEMG parameters in the same position window for each contraction cycle, a cyclostationary technique can be used (Bonato et al. 1996; Karlsson et al. 2003).
Unfortunately, static or standardized dynamic contractions are not common in daily life activities. In the fields of rehabilitation medicine, sports medicine and ergonomics, it is more natural for the subject to perform functional tasks similar to daily life activities or movements. Therefore, more advanced signal processing methods are needed to characterize non-stationary sEMG signals both in the time and frequency domains. A two-dimensional representation is needed, i.e. a time–frequency representation (TFR). Here, we briefly summarize these methods and refer to Hlawatsch & Boudreaux-Bartels (1992) for details.

(i) The short-time FT

The time resolution can be improved, compared with the epoch-dividing method, by sliding a window, \( w(t) \), along the sEMG and taking the FT for each local stationary segment. The short-time FT (STFT) for a given signal \( x(t) \) can be expressed as

\[
\text{STFT}(t, \omega) = \int w(\tau - t)x(\tau)e^{-j\omega \tau} \, d\tau, \quad \omega = 2\pi f. \tag{5.1}
\]

By performing FTs on a sequence of sufficiently short and equal time intervals (windows) of the signal, the time–frequency information is preserved. However, the precision of the information is limited and determined by the size of the window. A small window leads to high resolution in time but poor frequency resolution, whereas a large window gives the opposite. This is due to the fact that time and frequency resolutions (\( \Delta t \) and \( \Delta \omega \)) are dependent, and have to satisfy the uncertainty inequality, \( \Delta t \Delta \omega \geq 1/2 \).

(ii) Bilinear time–frequency representations

An alternative approach to linear TFRs (STFT and wavelets) is the popular Wigner–Ville distribution (WVD), defined as

\[
\text{WVD}(t, \omega) = \frac{1}{2\pi} \int x\left( t + \frac{\tau}{2} \right)x\left( t - \frac{\tau}{2} \right)e^{-j\omega \tau} \, d\tau, \tag{5.2}
\]

which is a bilinear TFR that has several desirable properties. For example, it is always real-valued, and time and frequency marginal properties, as well as time and frequency shifts to the signal, are preserved. In general, for a digitized \( n \)-component signal, the WVD has \( n \) auto-components and \( n(n-1)/2 \) interference terms. These interference terms (or cross-terms) are independent of the time–frequency distance between the two signal terms. In surface electromyography applications, when analysing multi-component non-stationary signals, the interference terms may overlap the auto-components. Adding a two-dimensional kernel function for linear (low-pass) filtering into the integral suppresses the interference terms. However, this also reduces the resolution. Therefore, it is important to find a kernel that preserves resolution and reduces the interference terms with respect to the analysed signal. For example, Bonato et al. (1996) used an exponential kernel proposed by Choi & Williams (1989) in applications for analysing the sEMG during dynamic contractions.
(iii) The continuous wavelet transform

Unlike the STFT that uses equal length analysis windows, the wavelet transform (WT) uses basis wavelet functions that have time widths adapted to each frequency band, i.e. the window is narrow for high frequencies and wider for low frequencies. Therefore, the WT acts as a mathematical microscope in which different parts of the signal can be observed by adjusting the focus. The continuous WT (CWT) is defined as

$$\text{CWT}(a, b) = \int x(t)\psi_{a,b}(t) \, dt,$$

(5.3)

where $a \in \mathbb{R}^+$ represents the scale parameter; $b \in \mathbb{R}$ represents the translation parameter; and $\psi_{a,b}(t)$ (the asterisk represents the complex conjugate) is obtained by scaling or dilation of the prototype or mother wavelet $\psi(t)$ at translation $b$ and scale $a$,

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right).$$

(5.4)

The wavelet is a smooth oscillating function with good localization both in time and frequency. One important property is that the wavelet integrates to zero, i.e. $\int \psi(t) \, dt = 0$. Thus, it oscillates as the name suggests. Good time and frequency localization implies compactly supported functions. This means that the wavelets decay in both the time and frequency domains, i.e. they actually become zero outside some interval. The wavelet’s smoothness (order of differentiability) can be chosen to satisfy conditions for a particular application. For muscle fatigue, see Karlsson et al. (2000) and the example in figure 2. A scale of $a > 1$ dilates the wavelet, while $a < 1$ compresses it, which is useful for the analysis of low- or high-frequency components, respectively. The factor $1/\sqrt{a}$ in equation (5.4) is introduced to guarantee energy preservation at every scale.

This time-scale expression has an almost equivalent time–frequency expression, since wavelets, e.g. the Morlet wavelet, are well localized around a non-zero frequency $f_0$ at scale $a = 1$ (i.e. the mother wavelet).

The representations of the methods presented in this section are highly redundant. However, sampling the $a$ and $b$ coefficients in the CWT on a dyadic grid leads to the discrete WT (DWT). This non-redundant multi-resolution representation is typically implemented using digital filter banks. In addition, a generalization of the DWT, the wavelet packets, in which arbitrary time-scale resolution can be chosen according to the signal, has also been applied to the sEMG (Karlsson et al. 1999). Similarly, von Tscharner (2000) adapted wavelet functions to achieve specified frequency resolutions.

(b) Spectral descriptors

Mean frequency (MNF) and median frequency (MDF; see Knaflitz & Bonato 1999) of the PSD are commonly used in the literature as spectral indicators. If the spectrum changes with time, the instantaneous mean frequency (IMNF) is preferable over the MNF. It can be calculated as

$$\text{IMNF}(t) = \int_{f_0}^{f_1} f \cdot P(t, f) \, df / \int_{f_0}^{f_1} P(t, f) \, df,$$

(5.5)

Phil. Trans. R. Soc. A (2009)
where $f_L$ is the lowest frequency and $f_H$ is the highest frequency in the bandwidth of the calculated TFR. Especially when the SNR is low, the choice of the bandwidth can affect the result. Therefore, a technique to find the maximum frequency of the sEMG spectra and to use that frequency as the integration limit was proposed by Knaflitz & Bonato (1999) and later improved by Östlund et al. (2004; figure 2).

### 6. Muscle fibre CV

Since CV depends on the intrinsic muscle fibre properties and their metabolic environment, CV is one of the most basic physiological parameters that can be estimated with surface electromyography. More specifically, the CV is related to, and thereby provides information about, muscle fibre cross-sectional area (Blijham et al. 2006), muscle fibre type (Kupa et al. 1995), MU recruitment (Houtman et al. 2003) and local muscle fatigue (Arendt-Nielsen & Zwarts 1989). Moreover, altered CV is an indication for some neuromuscular diseases (Blijham et al. 2008). As a consequence, sEMG signals are often collected to obtain information about CV. In contrast to the indirect estimation of
CV through spectral analysis using single-channel surface electromyography, CV can be directly measured with multiple electrodes above one muscle. The CV estimated using the sEMG is the average CV of all active MUs in the detection volume, with larger MUs contributing more to the average. However, in general, the CV of single or populations of MUs can be obtained by the same methods that give single MUAPs from the interference sEMG, e.g. some decomposition techniques (see §7) or the method described by Holtermann et al. (2008b; figure 3).

**Two-channel techniques**

The general approach is to estimate the time delay, $\Delta t$, between two signals recorded by two electrodes located at a distance $\Delta d$ apart, along the fibre orientation, and then $\text{CV} = \Delta d/\Delta t$. However, in practice, the signals are not identically delayed waveforms. In fact, they are dependent on many factors, such as the CV distribution of the active MUs, volume conduction inhomogeneities and generation and extinction effects on the MUAPs. This has led to many different methods to estimate the delay.

*Phil. Trans. R. Soc. A (2009)*
A basic approach to estimate the delay is through the maximum of the cross-correlation function,
\[
\max_{\Delta t} \sum_n x_1[n + \Delta t]x_2[n]. \tag{6.1}
\]

Owing to the limited time resolution (approx. 1 ms in the sEMG), interpolation techniques are required. A comparison of some of the above techniques can be found in Farina & Merletti (2004).

(b) Multi-channel techniques

Methods using electrode arrays (one-dimensional), or arrays within two-dimensional electrode grids, are reported to perform well when aligned with fibre direction, even in estimating CVs of single MUs (Farina et al. 2001; Houtman et al. 2003). In particular, the method proposed by Farina et al. (2001) has demonstrated high performance. Their approach is based on maximum-likelihood estimation, where the optimal delay \( \Delta t \) can be found as
\[
\min_{\Delta t} \sum_n \left( x_k[n] - \frac{1}{K-1} \sum_{m,m \neq k} x_m[n + (m-k)\Delta t] \right)^2, \tag{6.2}
\]
which is a modified beam-forming procedure, minimizing the mean square error between the signal \( x_k \) and the average of delayed versions of the \( K \) other synchronized signals \( x_m \).

Although the effect on CV estimation for slight inclination angles between detection system and fibres is small in theory, electrode placement in alignment with fibre direction is time consuming and many muscles have regions with fibres with different angles, and fibre angles can vary with contraction level.

This limitation was addressed by Grönlund et al. (2005b) using a spatio-temporal approach based on two-dimensional high-density sEMG recordings. High-density sEMG recordings using a two-dimensional electrode grid contain, in principle, information about muscle fibre orientation as well as the CV.

The method is based on three steps: (i) detection of trajectories of individual MUAPs propagating along the muscle fibres as recorded on the skin surface (Grönlund et al. 2005b), (ii) estimation of their corresponding CVs (Grönlund et al. 2005b), and (iii) separation of CV estimates recorded at different locations on the skin’s surface (Holtermann et al. 2008b). The method is described briefly here. A leading peak of a unidirectional propagating MUAP can be described as the set of positions in the coordinates of the electrode grid (row \( r \) and column \( c \)). Such MUAP trajectories are detected using an optimization algorithm based on running time window maximization of a cost function. Estimation of CV is accomplished using a spatio-temporal delay by fitting a spatio-temporal model, \( F \), of a propagating MUAP to the signals, \( X_{r,c} \), recorded within a time window,
\[
\min_{\Delta c} \sum_r \sum_c \sum_m \left| X_{r,c}[m] - F_{r+\Delta r,c+\Delta c}[m + \Delta t] \right|. \tag{6.3}
\]
The optimal delay $\Delta t$ and propagation components along and transverse to the electrode grid orientation, $\Delta r$ and $\Delta c$, can now be used to estimate the CV as

$$CV = \left(\frac{1}{\Delta t}\right)(\Delta r^2 + \Delta c^2)^{1/2}.$$  \hspace{1cm} (6.4)

In addition to the CV, the fitting procedure also gives information on initial spatial position $(r_0, c_0)$, shape characteristics and muscle fibre architecture with respect to the electrode grid of a propagating MUAP. In vivo fibre architecture using this technique was demonstrated by Holtermann et al. (2008b). Using the density distribution of estimates from CV and $c_0$ (figure 3), the CV of single (or, possibly, populations) of active MUs detected at different locations perpendicular to the fibre orientation can be separated. The density distribution can be estimated using two-dimensional kernel density estimation.

Another method that focuses on minimizing the influence of generation and extinction effects, and changes in waveform during propagation, was presented by Farina & Merletti (2003). This method is based on the temporal filtering of two sEMG signals recorded with adaptive spatial filters.

7. Single MU investigation (decomposition)

In clinical diagnostics, the activity and characteristics of individual MUs are commonly investigated. With information about single MU firing instances, MUAP shape and amplitude distribution, many physiological properties mentioned in the previous sections can be disentangled (Zwarts & Stegeman 2003). This gives information about motor control and muscle physiology in healthy and pathological conditions. Owing to the limited spatial selectivity of surface electromyography, needle electromyography is usually applied. However, surface electromyography has some advantages over needle electromyography, e.g. it is less painful, can be obtained during dynamic contractions and during daily life activities. In addition, it allows simultaneous recording at multiple sites, such that spatial aspects of MU activation can be investigated (Zwarts & Stegeman 2003).

The objective with decomposition is to separate the interference signal into individual MUAP trains. This is normally performed with invasive electromyography, where needles or wires are used to obtain the recordings. The decomposition is often divided into three main steps. The first step is a detection step, then there is a classification step (often using template matching) and, finally, there is a post-processing step, where, for example, superpositioned MUAPs are resolved. The processes are often similar for the sEMG, but there is more focus on the post-processing step and more a priori information is applied (e.g. De Luca et al. 2006; Kleine et al. 2007). For example, the a priori information about normal firing rates is often used. Of course, this has the drawback that normal firing patterns are preferred.

In blind source separation, which is the problem of finding sources by using their mixtures, independent component analysis (ICA) is now often used. ICA uses the central limit theorem ‘backwards’. The central limit theorem states that the summation of a large number of independent variables tends towards a normally distributed variable, regardless of the distribution of the original variables. To find the source signals, the ‘non-Gaussianity’ is thus maximized. In this context, the calculation is very similar to projection pursuit. However, in ICA, the number of sources is often known, while projection pursuit is used for
exploratory analysis of multivariate data. Maximizing the non-Gaussianity can be achieved in several different ways. A popular algorithm to perform ICA is the FastICA algorithm, which maximizes an approximation of negentropy (using contrast functions). There are also other algorithms, e.g. JADE (fourth-order cumulants), MILCA (based on mutual information), Kernel ICA (uses contrast functions) and RADICAL (uses an entropy estimator).

The ICA approach, which could potentially perform the decomposition in one step, has been tried for the sEMG (Nakamura et al. 2004). Unfortunately, the ICA technique does not work as it is, but could perhaps be a valuable part of a decomposition method. Holobar & Zazula (2004) used a multiple-input and multiple-output approach (as in ICA) to find different MUAP trains, but they used only second-order information. Unfortunately, their proposed procedure is quite sensitive to noise. Higher order statistics, which are used in ICA, have been used for decomposition of the sEMG. For example, the filtering approach proposed by Östlund et al. (2006) could be used for decomposition if different spatial locations are used to obtain different single MUAP trains (figure 4). The difference compared with ICA is that, in an ICA application, the different sources are found from the spatial information as different linear combinations of the same channels. The adaptive filter that was proposed by Östlund et al. (2006) finds different sources by using different spatial locations. Furthermore, the adaptive filter can use both the spatial and temporal information of the data—ordinary ICA only uses the spatial information.

The drawback with the more advanced techniques is that the shapes of the MUAPs are often assumed to be equal for every firing, which, of course, is not true. This can partly be solved by using the methods on short segments (assumed stationary) that later can be combined to a longer signal. However, the decomposition of the sEMG, even at low-force levels, is still far from routine work.

8. MU synchronization

Nearly simultaneous discharge of different MUAPs, called MU synchronization, is observed during fatigue (Holtermann et al. in press), in trained individuals (Semmler 2002) and in patients with central lesions (Datta et al. 1991). Because
MU synchronization reflects connections within the central nervous system, it may provide unique insights into functional neurophysiological aspects during different voluntary tasks and pathological conditions.

The most commonly used technique to quantify MU synchronization uses needle electromyography and histograms of the cross-correlation of pairs of single MUAPs. However, only a very small MU population of the active muscle is typically examined when estimating MU synchronization by needle electromyography. The representativeness of this small population and the physiological implications of the results have thus been questioned.

Several methods to quantify or assess MU synchronization from a larger population of active MUs have been proposed using surface electromyography. They include, among others, spike shape analysis (Gabriel et al. 2007), mean frequency (Kleine et al. 2001), and full-wave rectification and frequency analysis (Halliday et al. 1999). While the information about MU synchronization is dependent on central physiological factors, the sEMG is also influenced by peripheral factors such as MUAP shape and volume conduction properties. Therefore, evaluation using simulations is of great importance. Only two of the methods in the literature have been thoroughly evaluated in terms of factors influencing the sEMG (Farina et al. 2002; Grönlund et al. in press), and they are reviewed here.

A method based on the bipolar sEMG is the recurrent quantification analysis (RQA) approach (Farina et al. 2002) and is based on properties of the phase space (embedding) and thresholding. The MU synchronization is subsequently quantified using the per cent determinism. This gives a measure of repetitive structures within the signal. Although the method has demonstrated sensitivity to MU synchronization, it has a high dependency on the average CV. This is a major limitation since MU synchronization level and CV often change in parallel. In addition, all techniques using the bipolar montage gain information on activity from the most superficial MUs within the muscle, and therefore the technique is believed to be limited in terms of the examined MU population size.

The above limitations were reduced in the approach presented by Grönlund et al. (in press). The method is based on the asymmetry of monopolarly recorded MUAPs. With MU synchronization, the distribution of the monopolar sEMG will go from Gaussian (due to the central limit theorem) to a more skewed distribution. The descriptor is the skewness statistic used on sub-band filtered monopolar sEMG signals (sub-band skewness). A pre-processing step is used to reduce the dependency on MUAP duration of CV. This is accomplished by band-pass filtering the signals using WTs and Mexican hat wavelets at an optimal scale. This wavelet is chosen based on its shape similarity when compared with the monopolar MUAP. If we normalize the signal $x$ (of length $N$ samples) to zero mean and unit variance, then

$$\text{skewness} (x) = (N - 1)^{1/2} \sum_{n} x[n]^3$$

(8.1)

can be used to quantify changes in the MU synchronization throughout a contraction.

9. Localization of the IZ

Knowing the location of the IZ is important when recording an ordinary bipolar sEMG, and therefore recommendations for electrode placement to avoid the IZ are proposed (e.g. Hermens et al. 1999). However, the large inter-individual
differences in IZ location limit the application of the general recommendations. Knowledge of the IZ location is also important in other circumstances, e.g. for injection of botulinum toxin, and as a pre-processing step for some multi-channel sEMG analyses. Only two automatic methods for IZ localization have been proposed. One defines the location of the IZ as the bipolar channel with the lowest RMS (Rainoldi et al. 2004). This method is fast and simple, but unfortunately quite unstable. Another method is based on optical flow (Östlund et al. 2007). An optical flow field is a vector field that describes the movement between two images.

The optical flow field is obtained by assuming that the intensities in two consecutive images are the same, but they are translated by a small distance. That is, there exist $dx$, $dy$ and $dt$, such that

$$I(x + dx, y + dy, t + dt) = I(x, y, t),$$

(9.1)

where $I(x, y, t)$ is the intensity level of a dynamic image at position $(x, y)$ at time $t$. If the left-hand side of equation (9.1) is written as a first-order Taylor series (neglecting higher order terms), the optical flow constraint equation is obtained as

$$-I_t = I_x u + I_y v.$$  

(9.2)

Here, $u$ and $v$ are the velocities $dx/dt$ and $dy/dt$, respectively, and $I_x$, $I_y$ and $I_t$ are the partial derivatives of $I$. Since the velocity components ($u$ and $v$) are not uniquely defined from equation (9.2), Horn & Schunck (1981) introduced a smoothness constraint in their optical flow algorithm. They assumed that the velocity field changes slowly in a given neighbourhood. Since multi-channel surface electromyography can be seen as images describing the potential distribution over the skin, the optical flow can illustrate movements of the potentials on the skin. The IZ is then found as the origin of the flow (figure 3).

### 10. Final remarks

In this paper, only a sample of the processing methods in surface electromyography has been presented. The aim was to gain insight into neuromuscular physiology. The numerous processing solutions represent a challenge, but also an opportunity, for attaining non-invasive, valid and reliable measures of neuromuscular physiology during voluntary contractions and pathology. Great progress has been made, with promising results, in the processing methods in the relatively novel multi-channel sEMG technique (Zwarts & Stegeman 2003). This field represents an especially promising future for the application of the sEMG in clinical settings. However, the processing methods need to be standardized and automated to be applied by average users.

This work was supported by the European Union Regional Development Fund and the Swedish Research Council (K2004-27XX-15053-01A and 621-2007-3959). This project was performed within the Centre for Biomedical Engineering and Physics, Umeå University, Sweden.
References


Phil. Trans. R. Soc. A (2009)


