The extraction of neural strategies from the surface EMG: an update

Dario Farina,1 Roberto Merletti,2 and Roger M. Enoka3

1Department of Neurorehabilitation Engineering, Bernstein Focus Neurotechnology Göttingen, Bernstein Center for Computational Neuroscience, University Medical Center Göttingen, Georg-August University, Göttingen, Germany; 2Laboratory for Engineering of the Neuromuscular System, Department of Electronics and Telecommunications, Politecnico di Torino, Turin, Italy; and 3Department of Integrative Physiology, University of Colorado Boulder, Colorado

Submitted 19 February 2014; accepted in final form 21 September 2014

Farina D, Merletti R, Enoka RM. The extraction of neural strategies from the surface EMG: an update. J Appl Physiol 117: 1215–1230, 2014. First published October 2, 2014; doi:10.1152/japplphysiol.00162.2014.—A surface EMG signal represents the linear transformation of motor neuron discharge times by the compound action potentials of the innervated muscle fibers and is often used as a source of information about neural activation of muscle. However, retrieving the embedded neural code from a surface EMG signal is extremely challenging. Most studies use indirect approaches in which selected features of the signal are interpreted as indicating certain characteristics of the neural code. These indirect associations are constrained by limitations that have been detailed previously (Farina D, Merletti R, Enoka RM, J Appl Physiol 96: 1486–1495, 2004) and are generally difficult to overcome. In an update on these issues, the current review extends the discussion to EMG-based coherence methods for assessing neural connectivity. We focus first on EMG amplitude cancellation, which intrinsically limits the association between EMG amplitude and the intensity of the neural activation and then discuss the limitations of coherence methods (EEG-EMG, EMG-EMG) as a way to assess the strength of the transmission of synaptic inputs into trains of motor unit action potentials. The debated influence of rectification on EMG spectral analysis and coherence measures is also discussed. Alternatively, there have been a number of attempts to identify the neural information directly by decomposing surface EMG signals into the discharge times of motor unit action potentials. The application of this approach is extremely powerful, but validation remains a central issue.

surface EMG: motor unit; motor neuron; neural drive to muscle; amplitude cancellation; coherence; rectification; spectral analysis; decomposition

THE SURFACE EMG SIGNAL REPRESENTS the electrical activity generated in muscle fibers in response to the activation provided by innervating motor neurons. The information contained in the EMG signal comprises a mixture of details about the synaptic inputs received by the motor neurons and the muscle fiber electrical properties, as indicated by the following description of a surface EMG signal:

\[ s(t) = \sum_{i=1}^{M} \sum_{j=-\infty}^{+\infty} \varphi_i(t) \delta(t - t_{ij}) \]  

(1)

where \( M \) is the number of active motor units, \( s(t) \) is the recorded EMG signal, \( t_{ij} \) is the j-th time at which the i-th motor unit discharges action potentials, \( \varphi_i(t) \) is the shape of the action potential for the i-th motor unit, and \( \delta(t - t_{ij}) \) is the delta function representing a motor unit discharge at time \( t_{ij} \).

In the model depicted by Eq. 1, which for simplicity does not include additive noise, the discharge times \( t_{ij} \) of the motor neurons are determined by the convergence of synaptic inputs from multiple sources within the central nervous system. The set of discharge times for the activated motor neurons, often referred to as the neural drive to muscle, contains information about a variety of synaptic inputs received by the motor neurons. These inputs are characterized and can be identified by their frequency bandwidths.

The influence of the electrical properties of the muscle fiber membranes on the EMG signal is expressed by the shapes of the motor unit action potentials \( \varphi_i(t) \) (Eq. 1). The EMG signal is often analyzed as an interference signal and mathematical descriptors provide information about the global characteristics of the signal. The generative model of the EMG in Eq. 1 indicates that it is not possible to extract physiological information from an interference EMG signal that is solely related either to the neural drive [\( \delta(t - t_{ij}) \)] or to the muscle properties [\( \varphi_i(t) \)]. The measurement of EMG amplitude, for example, is frequently used to infer the intensity of the neural drive to the muscle even though it also depends on the properties of the muscle fiber membranes. Because the distribution of the shapes of action potentials varies across conditions, the amplitude-based estimate of neural drive to muscle contains some uncertainty. Similarly, the estimation of mean muscle fiber conduction velocity (33), a muscle fiber property, from the interfer-
ence EMG signal is influenced by the discharge characteristics of the active motor units (31, 39, 77, 95), which are neuronal properties.

In general, the distribution of the energies and shapes of the motor unit action potentials will compromise any method that attempts to extract information about the neural drive from an interference EMG signal. This constraint limits the use of the surface EMG for assessing the neural information because the distribution of motor unit action potentials depends on factors that cannot be controlled or determined in experimental conditions. These factors include the thickness of subcutaneous tissues, the spatial distribution of muscle fibers, the distribution of muscle fiber conduction velocities, and the electrode location (13, 28, 31, 91).

In our previous review on this topic (31), the main emphasis was on amplitude cancellation and on the use of the frequency analysis of the EMG for studying motor unit recruitment strategies. Many of the problems we identified have been progressively acknowledged over the last 10 yr, sometimes after further discussion (34, 35, 102, 103). The present review provides an update on the extraction of information about the neural drive to muscle from EMG signals. As in our previous article (31), the current review focuses on selected issues and does not provide an extensive review of the methods for information extraction and processing from the EMG signal.

Indirect Analysis of Neural Strategies from the Interference EMG

As discussed in our previous review (31), the amplitude and spectral properties of a surface EMG signal have often been used to provide indirect measures of neural activation, such as details about the intensity of the neural drive or motor unit recruitment strategies. In the following, we provide an update on these practices and then broaden the discussion to the use of coherence methods for studying neural connectivity.

Amplitude cancellation. The most basic approach to extract information from an EMG signal is to estimate its amplitude. Because the EMG signal has random components, its amplitude cannot be defined on the basis of specific waveform features, such as the peak value, but rather needs statistical estimators. In signal processing terminology, the EMG is a colored stochastic process, a random signal with limited bandwidth, and its amplitude is its standard deviation, which is the square root of the power (8). Given the EMG signal model of Eq. 1 and assuming the trains of action potentials discharged by the active motor neurons are not correlated, the standard deviation (amplitude) \( \sigma_{EMG} \) of the EMG signal is theoretically obtained with the following expression:

\[
\sigma_{EMG} = \sqrt{\sum_{i=1}^{M} P_i} = \sqrt{\sum_{i=1}^{M} DR_i E_i}
\]  

where \( \sigma_i \) and \( P_i \) are the amplitude and power for the train of the action potentials discharged by the \( i \)-th motor unit \( \sigma_i^2 = P_i \), \( DR_i \) is the average discharge rate of the \( i \)-th motor unit, and \( E_i \) is the energy of the action potential of the \( i \)-th motor unit. In Eq. 2, \( DR_i \) represents the influence of the neural drive and \( E_i \) denotes the contribution of action potential shapes to the estimate of signal amplitude. Equation 2 indicates that the EMG amplitude \( I \) increases monotonically when either the discharge rates or the number of active motor units increase; thus it is directly associated with the neural drive to muscle; and 2) depends on both the neural drive (discharge rates, \( DR_i \)) and the properties (energy, \( E_i \)) of the action potentials, without the possibility of distinguishing between the two effects. Normalization methods, such as dividing an EMG signal by the amplitude recorded at the maximal force (73), attempt to reduce the influence of the distribution of action potential energies in the right-hand side of Eq. 2. However, Eq. 2 indicates that complete elimination of the influence of action potential energies is only possible when all motor unit action potentials have the same energy.

Equation 2 also indicates that the estimation of EMG amplitude from an interference signal is less than that obtained by summing the amplitudes of the individual motor unit action potentials, which is a phenomenon referred to as amplitude cancellation (11, 61). Formally, amplitude cancellation is due to the following inequality:

\[
\sigma_{EMG} = \sqrt{\sum_{i=1}^{M} DR_i E_i} \leq \sum_{i=1}^{M} \sigma_i
\]

The amplitude cancellation term corresponds to the difference between the sum of the amplitudes of the action potential trains \( \sigma_i \) and the amplitude of the EMG, \( \sigma_{EMG} \). Because the difference between the two estimates of EMG amplitude depends on the number of activated motor units (Eq. 3), amplitude cancellation increases monotonically with an increase in the neural drive to muscle. Moreover, amplitude cancellation is greatest when the action potential shapes are the same for all motor units (36), and in this instance EMG amplitude is proportional to the square root of the neural drive to muscle (Eq. 3).

Amplitude cancellation differs among motor units and depends on the ratio between the amplitude of the motor unit action potential and the amplitude of the EMG signal (36). For example, motor unit action potentials with a root mean square amplitude less than that for the interference EMG are cancelled by >60% (see Fig. 2 in Ref. 36). Because surface-recorded action potentials of low-threshold motor units tend to be smaller, amplitude cancellation generally reduces the contribution of low-threshold motor units to the estimate of EMG amplitude derived from an interference signal. Consequently, estimates of surface EMG amplitude can be relatively insensitive to changes in the activity of low-threshold motor units. For example, Mottram et al. (78) found that the discharge rate of low-threshold motor units decreased more rapidly during one type of fatiguing contraction compared with another type, whereas the change in the amplitude of the surface EMG was similar during the two tasks. Due to this effect, indexes of neural drive to the muscle based on EMG amplitude, such as the concept of neuromuscular efficiency (18), largely reflect the activity of high-threshold motor units.

The influence of amplitude cancellation on the representation of motor unit activity in the surface EMG is shown in Fig. 1. In this simulation, 300 motor units were active during a sustained muscle contraction in which the discharge rates of 50 low-threshold motor units decreased progressively over time during a fatiguing contraction to half of their initial values. Despite the decrease in discharge rate for 17% of the activated motor units, the estimated EMG amplitude remained relatively constant during the simulated 5-s contraction due to the small influence of these motor units on the overall signal amplitude.
This example illustrates the tendency of amplitude cancellation to reduce the relative contribution of individual or groups of motor units to estimates of EMG amplitude.

Moreover, rectification of an EMG signal does not eliminate the influence of amplitude cancellation (11, 61). A rectified EMG signal has the same amount of amplitude cancellation as an interference EMG signal. Consequently, the amplitude of a spike-triggered average derived from a rectified EMG signal depends on the amount of amplitude cancellation (36, 62).

In contrast to the effect of amplitude cancellation on estimates of EMG amplitude, the square of each term in Eq. 2 indicates that the power of an interference EMG signal is the sum of the powers of the trains of motor unit action potentials (provided the trains of motor unit action potentials are uncorrelated). Amplitude cancellation indeed does not influence the power spectrum (i.e., the distribution of signal power across frequencies) of an interference EMG signal: the power spectra of uncorrelated trains of motor unit action potentials add linearly to generate the power spectrum of the interference EMG signal. Similarly, EMG rectification does not influence the total signal power, which is obtained by integrating the squared signal and dividing by the time interval of the integral. Thus the power of the rectified EMG is equal to the power of the interference EMG. Nonetheless, the distribution of power across frequencies (power spectrum) changes with rectification because it is a nonlinear operator and is influenced by cancellation: the power spectrum of a rectified EMG signal is not equal to the sum of the power spectra of the rectified trains of motor unit action potentials (41). Due to this effect, the spectral peak of an interference EMG signal corresponding to the average motor unit discharge rate is not consistently augmented by rectification (31). Claims to the contrary (79) are based on neglecting the cancellation term in the power spectrum of the rectified EMG (see also EMG rectification and coherence analysis).

These observations lead to several conclusions, some of which were partially presented in our previous review (31) and discussed in subsequent work (36, 41, 59, 61, 62). First, EMG amplitude derived from a global measure of muscle activity underestimates the sum of the amplitudes of the trains of motor unit action potentials, and the magnitude of the deficit cannot be predicted or compensated (61). Consequently, EMG amplitude provides only a crude estimate of the neural drive to muscle (31). Second, because amplitude cancellation increases monotonically with the level of muscle activity, the intensity of the neural drive to muscle is the main determinant of the degree of amplitude cancellation (36, 61). Third, the amount of amplitude cancellation differs across motor units and is greatest for small surface-recorded action potentials, which usually are associated with low-threshold motor units (36). Fourth, the rectified EMG has the same degree of amplitude cancellation as an interference EMG, but amplitude cancellation only influences the power spectrum of a rectified EMG signal and not the interference EMG signal (41). The latter conclusion is critical for the use of signal rectification in coherence analysis (see EMG rectification and coherence analysis).

**EMG spectral analysis and motor unit type.** In our previous review (31), we criticized the use of spectral properties extracted from a surface EMG signal for inferring details about the recruitment of motor units. This practice is based on the observation that muscle fiber conduction velocity, which sometimes is associated with different motor unit properties, scales the EMG power spectrum (69, 94): slower conduction velocities are associated with lower frequencies in the power spectrum. This approach is compromised, however, by the absence of a physiological distinction between purported motor unit types (25, 31, 34, 35, 100) and the influence of the intracellular action potential shape and of the volume conductor on the surface action potential power spectra (28, 30). For example, the difference between a surface action potential generated by a motor unit with relatively high conduction velocity (high threshold) and that generated by a motor unit with lower...
conduction velocity (lower threshold) may be compensated by a different average distance between the two motor units and the recording electrodes (volume conductor effect) (see Fig. 3 in Ref. 31). Following our previous review (31), there was a more detailed discussion of this issue (34, 35, 102, 103) and it is now recognized that these limitations render the approach meaningless for the vast majority of situations. The number of studies that have used this approach has decreased over the last 10 yr, although inappropriate interpretations of EMG power spectra are still being published (e.g., Ref. 89).

The use of EMG spectra to infer information about motor unit activity is even more compromised during sustained contractions when the neural drive and muscle properties vary over time. As an example of these complications, Fig. 2A shows the results of simulated ramp contractions performed before (left) and after (right) a fatiguing contraction with the model proposed by Dideriksen et al. (19). The ramp contractions involved a gradual increase in the simulated force and then a gradual decrease (Fig. 2B). The simulated EMG signals were analyzed with a time-frequency method (spectrogram) (66). In this simulation, motor units were recruited and derecruited according to the size principle (44) during each ramp (Fig. 2C), which resulted in progressive increases and decreases in the amplitude of the interference EMG signal (Fig. 2D). The recruitment and derecruitment were also related to the conduction velocity of the innervated muscles fibers in that higher threshold motor units had greater conduction velocities (1). Moreover, conduction velocity decreased during the sustained contractions.

Fig. 2. Sensitivity of surface EMG power spectral frequencies to motor unit activity. An integrated model developed by Dideriksen et al. (19) was used to simulate a fatiguing contraction. The model includes the modulation of motor unit activity and the electrophysiological and contractile properties of the muscle fibers by feedback derived from metabolic factors. The simulated protocol included an up-down ramp contraction to 30% of maximal voluntary contraction (MVC) force, a 140-s contraction sustained at 30% MVC force, and a final ramp contraction similar to the 1st one (A). The model predicted a decrease in peak force for the 2nd ramp contraction after the fatiguing contraction (B). The motor units were recruited and derecruited during the up and down phases of the ramp contractions. Discharge times are shown for every 10th motor unit (C). There was more motor unit activity during the 2nd ramp contraction. A bipolar surface EMG signal (D) was simulated with the volume conductor model proposed in Farina et al. (32), with parameters described by Keenan et al. (63). The instantaneous mean frequency (IMF) was computed from the spectrogram of the surface EMG signals simulated during the 2 ramp contractions (E). There was no association between the estimates of instantaneous mean frequency and the recruitment and derecruitment of motor units during the ramp contractions. Moreover, the estimated mean frequency was less during the 2nd ramp despite greater motor unit activity, indicating that the approach cannot detect the recruitment of motor units during fatiguing contractions.
contraction and there was a greater amount of motor unit activity during the ramp contraction after the fatiguing contraction (Fig. 2, C and D). Despite significant modulation of motor unit activity during the ramp contractions, the instantaneous mean frequency of the EMG remained relatively constant (Fig. 2E). Moreover, the instantaneous mean frequency was lower after the fatiguing contraction despite the greater amount of motor unit activity.

Neural connectivity. EMG signals are also often analyzed to extract information about the synaptic input received by the motor neurons and thereby to derive an indirect measure of the excitation signal received by motor neurons. These approaches are frequently used to quantify the strength of connections between different parts of the nervous system. For example, the linear correlation in the frequency domain (coherence) between EEG (or other cortical signals) and concurrently recorded EMG has been extensively used to quantify the strength of corticospinal input to motor neurons (2, 9, 47, 92). The connection is indicated by the value of EEG-EMG coherence in the beta band (15–30 Hz). Similarly, the coherence between EMG signals recorded from two muscles has been used to assess the strength and frequency of the common synaptic input received by the motor neurons innervating the two muscles (6, 10, 24). The same approach has been applied to EMG signals measured from two locations over the same muscle to estimate the common input to the motor neuron pool (48).

Because coherence methods are based on the assumption of a linear association between signals, it may seem inappropriate for them to be used to estimate the strength of association between signals derived from the nonlinear transformation of inputs into outputs by motor neurons. Subsequent sections indicate how this limitation is attenuated by the parallel distribution of common input to multiple motor neurons.

Linear transmission of common input by motor neurons. Oscillatory inputs arising from either cortical or peripheral sources are transmitted to the motor neuron output via a nonlinear transformation that distorts the input signal when it is processed by a motor neuron. Figure 3 shows this property by simulating neurons as perfect integrate-and-fire systems. When a simple sinusoidal input is transformed by a motor neuron into a train of motor unit action potentials (Fig. 3A), the output signal is modulated at both the frequency present in the input signal and other frequencies (harmonics) not present in the input signal. The additional harmonics in the output signal can be seen as noise with respect to the linear transmission of the input harmonic. The noise increases progressively when the signal is processed by a series of neurons in a pathway (Fig. 3A). The frequency and phase of the added harmonics differ for each motor neuron (80, 82). The frequencies added to the output signal by the nonlinearity of the motor neurons are equivalent to independent components across the motor neuron outputs. Because the cumulative output signal produced by a set of motor neurons, neural drive to the muscle, represents an average output signal of the motor neurons, the common components are augmented relative to the independent components derived from the nonlinearity (Fig. 3B). Transmission of an oscillatory signal by the central nervous system, therefore, is most effectively accomplished by distributing the input signal to sets of neurons.

The transmission of oscillations embedded in synaptic input into the neural drive to muscle is approximately linear when a sufficient number of motor neurons receive the input concurrently. Relatively few motor neurons need to receive the common input to achieve linear transmission of oscillations (82). Moreover, the averaging in the cumulative output train of action potentials enhances the common input against both the components introduced by the nonlinearity of the motor neurons (Fig. 3) and the components in the input to the motor neurons that are independent across motor neurons. The results of a simulation shown in Fig. 4 indicate how the number of motor neurons influences the transmission of the spectral characteristics of a common input signal into those of the cumulative output signal when both nonlinearities and independent input components are present. The progressive distribution of a common input signal is able to attenuate the influence of independent inputs and nonlinearity on the characteristics of the output signal (43). At low forces, however, there may be too few active motor neurons to ensure exact transmission of common input to the neural drive to muscle, which would influence the precision of the force generated by a muscle. For example, at forces less than ~10% of maximum, but not at higher forces, synaptic noise (independent input) can have an influence on force steadiness (20, 85).

The above considerations lead to the following fundamental observations when discussing the use of EMG for studying neural connectivity (42). First, the common input received by a set of motor neurons is transmitted linearly into the cumulative output (neural drive to the muscle), despite the nonlinearity of the transformation by individual motor neurons. Second, the inputs that are independent across motor neurons are also attenuated in the neural drive to muscle due to the averaging process. Third, the linearity in the transmission and the degree of attenuation of independent inputs depend on the number of motor neurons that receive the common input. Although these observations justify the use of coherence (linear) methods when investigating neural connectivity with EMG signals, the approach should be used with caution given the limitations described in the following sections.

EEG-EMG coherence and the strength of corticospinal input. Oscillations at ~20 Hz that are present in EEG signals in humans (9, 87) are coherent with the surface EMG activity during sustained contractions (47, 92). It has been suggested that this observation, so-called corticomuscular coherence, indicates the effective transmission of cortical oscillations along the corticospinal tract and through the monosynaptic connections between cortical neurons and spinal motor neurons into trains of motor unit action potentials (2). The amount of significant coherence between cortical and muscular activity is assumed to indicate the strength of the supraspinal contributions to the output generated by spinal motor neuron. Accordingly, abnormal corticomuscular coupling has been found in pathological conditions (46), such as stroke (26, 98) and tremor (93, 104), and in healthy subjects during fatiguing contractions (99), peripheral nerve stimulation (50), and manipulation of afferent pathways (88, 90).

The observation of significant EEG-EMG coherence is explained by the linear transmission of synaptic input received by an adequate number of motor neurons. However, the use of EEG-EMG coherence to quantify the strength of corticospinal input is constrained by at least two factors. First, the neural...
drive to muscle is filtered by the motor unit action potentials (Eq. 1) and their shapes will confound the estimated coherence. This influence can be readily appreciated by comparing the coherence estimated between an EEG signal and EMG signals recorded from different electrode locations over the same muscle, which comprised different action potential shapes (Fig. 5). The comparison of EEG-EMG coherence across conditions in which the action potential shapes change substantially, such as fatiguing contractions (21, 23), exercise, or differences in subcutaneous layer thickness (27, 28), should thus be performed with caution.

Second, the amount of EEG-EMG coherence can also be modulated by other sources of common input to motor neurons. Any common input, but not independent input, is transmitted through the active motor neurons. If the motor neurons receive a second source of common input, in addition to independent inputs and the common corticospinal input, its intensity will influence the estimate of corticomuscular coherence even when the relative proportion of corticospinal input does not vary (83). This type of modulation is referred to as decorrelation effect of secondary sources of common input (83). For example, a decrease in corticomuscular coherence during fatiguing contractions (106) can be partially explained by decorrelation due to changed (common) afferent input and does not necessarily indicate a decrease in the strength of corticospinal connection. The same decorrelation effect by secondary sources of common input will also be illustrated for EMG-EMG coherence.

**EMG-EMG coherence.** Coherence between EMG signals from different muscles indicates the presence of common synaptic input to the motor neuron pools of the two muscles. The presence of multiple common synaptic inputs attenuates the capacity of coherence to estimate the strength of common input from a single source, as discussed for EEG-EMG coherence. The decorrelating influence of a second common input is illustrated in Fig. 6. In this example, the motor neuron pools of two muscles receive a common cortical input in the beta band and each of them also receives afferent input from the muscle.
it innervates. The estimated coherence changes substantially when the afferent input of any of the two muscles varies relative to the other sources of input (Fig. 6C), despite the absence of change in the relative strength of the cortical common input to the two muscles. The same effect can be demonstrated for corticomuscular coherence when common afferent input is added to the corticospinal input (83).

In addition to the influence of a second source of common input, estimates of EMG-EMG coherence are confounded by the presence of EMG cross talk (12, 22, 29, 60), which is greatest for neighboring muscles. Methods to reduce the amount of cross talk and thereby improve estimates of EMG-EMG coherence have been proposed (60) but are based on the incorrect assumption that cross talk always increases the correlation between signals (see Ref. 31 for details). There are currently no analytical procedures that can eliminate cross talk in a recorded EMG signal.

Similar limitations apply when EMG-EMG coherence is used to assess the degree of common input to the same motor neuron pool (48, 86). This technique has been applied as a substitute for the coherence analysis between pairs of motor units within a muscle. Despite the limitations discussed, the fundamental flaw of this method is that groups of motor units from the same muscle will tend to generate the same cumulative train of motor unit action potentials. Indeed, for any given ratio between the powers of common and independent inputs to the motor neurons, the cumulative spike train of a group of motor neurons will average the common information and attenuate the independent components. Groups of motor units in the same muscle will thus tend to represent only their common input (43) (Fig. 4). Because surface EMG signals reflect the activity of many motor units, the coherence between EMG signals from the same muscle is relatively insensitive to the strength of common input with respect to independent input (level of synchronization) (84; see also Fig. 6 in Ref. 43). The neural information in the two EMG signals is quite similar and the variability in coherence values is mainly due to differences in the shapes of the action potentials.
EMG rectification and coherence analysis. In some conditions, it is not possible to detect oscillations in the discharge times of motor units in the surface EMG power spectrum or in a coherence analysis. This occurs when an input frequency is outside the bandwidth of the surface motor unit action potentials (41, 42). For example, Fig. 7 presents the case of an oscillatory input at 6 Hz, which is too low to be detected in the surface EMG signal spectrum. It is not possible to recover such low input frequencies with linear filtering of the surface EMG because these methods act as a gain and phase shift of each frequency, and frequencies with null power will still have null power for any applied gain. However, it is possible to retrieve the information on the input frequency with nonlinear methods under some conditions. Nonlinear transformations can introduce power at frequencies not present in the original signal. One nonlinear transformation that has received considerable attention in the surface EMG literature in relation to coherence analysis is rectification (49, 74, 75).

Rectification involves transforming negative values in a signal into equivalent positive values. Because rectification is considered to be an operation that augments the neural information of the signal relative to that associated with the shape of the action potentials, it is recommended as an initial step before performing coherence analysis (49). However, it is only possible to augment the oscillatory components of a neural input in the rectified EMG under some specific conditions.

The rectified EMG can be expressed as follows:

\[ s(t) = \sum_{i=1}^{M} \sum_{j=-\infty}^{+\infty} \varphi_i(t) \delta(t - t_i) + c(t) \]

where the rectified signal, \( s(t) \), is the sum of the recorded signal with rectified action potentials and the cancellation signal, \( c(t) \), which is equal to the difference between the sum of the trains of rectified motor unit action potentials and the rectified EMG signal (a minus sign can be integrated in this
The sum of the trains of rectified motor unit action potentials, which is indicated by the first term on the right-hand side of Eq. 4, is referred to as the no-cancellation signal (61). The no-cancellation signal has identical neural information as the interference EMG signal but with rectified motor unit action potentials. Comparison of Eqs. 4 and 1 indicates that the rectified EMG signal has the same structure as the interference EMG signal except that the motor unit action potentials are rectified and there is the addition of the noise term \( c(t) \), whose power depends on the degree of amplitude cancellation. The rectified EMG, therefore, represents a noisy version of the no-cancellation EMG (note that the noise is correlated to the signal in this definition). Thus there is no obvious reason why rectification should be expected to augment the neural information (discharge times) embedded in an interference EMG signal. Because oscillatory components in the neural input are filtered by rectified action potentials and are subject to an additional noise term, rectification can only augment the oscillatory inputs (neural information) if the energy of the rectified action potentials at the input frequency is greater than that of the recorded action potentials (as usually occurs for low frequencies) and if the noise introduced by the cancellation term is limited (41, 42). Moreover, there is also no obvious reason why rectification should be a preferred nonlinear transformation over many other possible transforms.

The difficulty to detect synaptic inputs to the motor neurons at relatively low frequencies in the power spectrum of the surface EMG signal is shown in Fig. 7A for an input oscillation at 6 Hz. Nonetheless, rectification of the EMG signal can recover the input information that was transmitted by the
involved motor neurons (Fig. 7A). Any nonlinear transformation that introduces power at low frequencies, specifically at 6 Hz in this example, will have a similar effect, as shown in Fig. 7A for the squaring and the logarithmic operators. The application of each of these nonlinear operators to the interference EMG signal is not equivalent to their application to each train of action potentials separately, unless there is no temporal overlap among action potentials. In all other conditions, a noise term, similar to that shown in Eq. 4, is introduced to characterize the difference between the application of the nonlinear operators to each action potential and to the interference signal. The specific noise term associated with each nonlinear transformation will depend on the intensity of muscle activation, as observed for rectification.

To influence of the noise term in extracting oscillatory inputs from the EMG, spectral representation is shown by simulating the same motor neuron pool and 6-Hz input (Fig. 7A) with an increased level of excitatory synaptic input to the motor neurons (Fig. 7B). In this instance, more motor neurons are recruited and discharge rates are greater, which increases both the amplitude of the interference EMG signal and the power of the noise term due to the nonlinearity of the operators (rectification, square, logarithmic). The increase in amplitude cancellation greatly reduces the capacity of nonlinear transformations to identify oscillations in the input signal (Fig. 7B).

The model shown in Eq. 4 allows prediction of the effectiveness of nonlinear EMG signal transformations to identify oscillatory synaptic inputs and explains several separate empirical observations. For example, Ward et al. (105) recently showed that rectification can indeed improve coherence analysis at low contraction forces. Moreover, Boonstra and Breakspear (7) noted that the uniformity in shapes of the action
potentials across motor units and the type of signal filtering had an influence on the effectiveness of EMG rectification in coherence analysis, which is explained by the influence of these factors on amplitude cancellation (36, 61).

The above considerations provide a clear view of the value and limits of rectification in EMG coherence analysis and can be summarized as follows (41). First, the rectified EMG is a noisy version of the no-cancellation EMG, where each action potential is rectified. Second, the noise term in the rectified EMG is the amplitude cancellation signal, as defined in Amplitude cancellation. The noise term for other nonlinear transformations has similar properties as amplitude cancellation and increases in magnitude with the intensity of muscle activity. Third, the spectral and coherence peaks associated with oscillatory input frequencies are amplified by nonlinear transformations, such as rectification, if the transformed action potentials have greater energy at the input frequency than the recorded action potentials and if the noise term introduced by the nonlinear operator is sufficiently small. Because this noise term always increases with the intensity of muscle activation, nonlinear transformations are only effective for coherence analysis at low contraction forces (Fig. 7).

Direct Analysis of Neural Strategies by Decomposing the EMG

Many of the limitations in the interpretation of the surface EMG are due to the signal containing information about the neural input (common and independent inputs) mixed with the shapes of the motor unit action potentials (Eq. 1). Separation of the two sources of information is challenging due to the absence of a priori information (51), such as the number of active motor units and the shapes of the action potentials. When the shapes of the action potentials discharged by the same unit differ less than those generated by other motor units, it is possible to distinguish the two sources of information. To verify this condition, Farina et al. (37) performed a simulation study to determine the proportion of motor units with unique surface action potentials; that is, action potentials that differed from those of all the other units in a representative sample. This approach does not depend on the method used to decompose the interference signal into its constituent trains of motor unit action potentials, but rather establishes limits for decomposition that are valid for any algorithm. The results indicated that only a small proportion of motor unit action potentials are unique when the recordings comprise a few channels of surface EMG signals. For example, <60% of the simulated motor units could be discriminated on the basis of the action potential shape with a recording system that comprises four channels (14). The remaining 40% of the simulated motor units had action potentials that were identical-within experimental variability—to those of at least one other motor unit. Although it may be possible to decompose the surface EMG signal from few recorded channels under some circumstances, the results do not generalize and the methods have only limited applicability.

The action potentials of the active motor units can only be distinguished with adequate spatial information, which requires many recording channels of the EMG signal. An increase in the number of channels in which each motor unit action potential is represented will increase the number of motor units that can be uniquely detected at the skin surface (37), rendering the decomposition challenge theoretically possible. To achieve this level of resolution, high-density surface EMG systems, with tens to hundreds of electrodes covering individual muscles, have been developed in several laboratories (4, 45, 56, 65, 67). In addition to their use for EMG decomposition, high-density electrode systems have also been used to map the electrical activity in muscle (64). This approach can provide robust force estimators from EMG amplitude (96), characterize the adjustments during fatiguing contractions (38, 97), and identify details about muscle architecture (67). In the present review, however, we focus only on the use of high-density EMG technology for decomposing the signal into individual trains of motor unit action potentials.

Among the methods developed to decompose a surface EMG signal, only the approach developed by Holobar and Zazula (51) and subsequently extended (54, 55) has been validated systematically. Decomposition accuracy has been reported as ~95% for a range of conditions and muscles, including high levels of motor unit synchronization (54), strong forces [up to 70% maximal voluntary contraction (MVC)] (55), complex muscle architectures (72), and pathological EMG patterns (68). Figure 8 provides an example of this approach by showing the decomposition of high-density surface EMG signals into discharge times for nine motor units. The decomposition of concurrently recorded intramuscular EMG signals is also reported and compared with the surface EMG decomposition. This comparison of discharge times for three motor units that were detected in both the surface and intramuscular signals provides a measure of accuracy of the decomposition results.

Validation of decomposition of surface EMG signals. Surface EMG recordings are more difficult to decompose into motor unit discharge times than intramuscular EMG signals due to the greater number of overlapping action potentials in the surface signal. The rate at which motor unit action potentials overlap in an EMG signal can be estimated as \((DR)^2 \cdot I \cdot (I - 1)\) (3), where \(DR\) is the average discharge rate [pulses per second (pps)] of the active motor units, \(d\) is the average duration (s) of the action potentials, and \(I\) is the number of motor units contributing to the signal. According to this estimate, an accurate surface EMG decomposition of 20 motor units discharging action potentials at 20 pps and with each surface-recorded action potential lasting 20 ms requires the resolution of ~3,000 overlaps each second. In comparison, a selective intramuscular signal that comprises 3 motor units discharging at 20 pps and with 5-ms action potentials has an overlap rate of only 12 action potentials per second. These estimates are typical of the problems encountered when decomposing surface and intramuscular EMG signals.

The amount of overlap in surface EMG signals poses a challenge not only for the decomposition but also for judging the validity of a decomposition algorithm. Figure 9A provides an example of this difficulty based on recordings of a surface EMG signal detected by a bipolar system with 5-mm interelectrode distance during a contraction of the tibialis anterior muscle in which the subject tracked a single motor unit based on a concurrently recorded intramuscular EMG. Despite a low contraction force (~2% MVC) and relatively high selectivity of the surface EMG electrode system (small interelectrode distance), the surface recording (top trace in Fig. 9A) shows a
high level of interference from which it is impossible to identify individual motor unit action potential trains by visual inspection. This uncertainty is underscored by comparing the surface EMG recording with the concurrently recorded intramuscular EMG (bottom trace in Fig. 9A). Due to the difficulty in the visual identification of single motor unit activities in the surface EMG, any set of automatically decomposed discharge times could be interpreted as correct. Contrary to the intramuscular EMG that can be decomposed manually by an expert and for which the manual decomposition is often the gold standard to assess accuracy, it is not possible to decompose surface EMG signals manually.

A convincing validation of surface EMG decomposition requires either that the method is applied to a sufficiently large number of test signals (53) or that it estimates the accuracy for each specific decomposed motor unit action potential train in a signal (signal-based accuracy) (55). In both cases, there is the need for valid performance indexes, although their validity may be counterintuitive. For example, Hu et al. (57) proposed intuitive criteria to judge the accuracy of an algorithm: 1) the averaged action potentials triggered by the decomposed trains of discharge times are highly correlated with the action potentials identified by the decomposition algorithm (but not if random variability is introduced in the triggers); and 2) the correlation between the original signal and the signal obtained by progressively adding the decomposed action potential trains is relatively high (but not after introducing variability in the identified discharge times). To assess the validity of these criteria, we implemented the approach in a de novo decomposition algorithm.

A single-channel surface EMG trace was decomposed by identifying the peak in the signal, extracting a segment of the signal around this peak, and detecting all other times when the extracted segment correlated with the signal above a fixed correlation threshold. The times when the correlation exceeded the threshold were assumed to indicate the discharge times of a single motor unit and detected by the algorithm. Moreover, the shape of the “motor unit action potential” associated with these discharge times was estimated by averaging the signal over the identified discharge times. The estimated motor unit action potential shape was finally subtracted from the signal at these discharge times was estimated by averaging the signal around this peak, and detecting all other times when the extracted segment correlated with the signal above a fixed correlation threshold. The times when the correlation exceeded the threshold were assumed to indicate the discharge times of a single motor unit and detected by the algorithm. Moreover, the shape of the “motor unit action potential” associated with these discharge times was estimated by averaging the signal over the identified discharge times. The estimated motor unit action potential shape was finally subtracted from the signal at these discharge times. These steps were applied iteratively to the residual signal until it had a power below a user-selected threshold. Such an algorithm will always extract “trains of discharge times” and produce corresponding “action potential shapes,” but the accuracy is poor in most cases.
Fig. 9. Validation of a new “decomposition” algorithm based on the criteria described by Hu et al. (57). A: surface and intramuscular EMG signals were recorded concurrently during a contraction at ~2% MVC force. The subject was provided with visual feedback on the discharge of a single motor unit from an intramuscular EMG recording (bottom trace). The motor unit with the greater amplitude action potentials was selected for the feedback signal. The algorithm described in the text identified the discharge times of several motor units in the surface EMG signal, which are indicated by the black circles. Although there was an association between the discharge of the target motor unit (triangles) and some of the discharge times detected by the algorithm, the rate of agreement was <15% even though the target motor unit had relatively high power in the surface EMG signal. B: despite the poor performance of the algorithm, spike-triggered averaging of the surface EMG signal from the discharge times identified by the new algorithm provide an action potential shape that was similar (correlation coefficient, 0.99) to that given as an output by the decomposition algorithm. This association disappeared for small perturbations (in the same range as in Ref. 57) of the discharge times, so that the correlation coefficient between the averaged potential with perturbed triggers and that provided by the algorithm decreased to only 0.27. C: progressive addition of the trains of decomposed action potentials (in order or in reverse order; filled and empty circles) produced correlation coefficients with the recorded EMG signal that increased monotonically, in contrast to the addition of perturbed discharge times (empty triangles) or random assignment of the decomposed action potentials to the trains of discharge times (black triangles). The results described in B and C have been interpreted as indicating the valid decomposition of a surface EMG signal (57), despite its inadequate performance as shown in A.

This approach was tested on the signal in Fig. 9A, the greatest correspondence between discharge times detected by the surface and intramuscular EMG signals was indeed <15%. Nonetheless, when tested by the spike-triggered averaging method (Fig. 9B) or by the spike-train correlation method (Fig. 9C), as in Hu et al. (57), the algorithm appeared to be accurate (high correlation in the spike-triggered potentials and in the sum of extracted spike trains). Moreover, there are an infinite number of algorithms that would satisfy the criteria established by Hu et al. (57) without being accurate. This example is not meant to imply that any algorithm tested with this approach is necessarily inaccurate (as for our de novo “decomposition algorithm”) but rather that the proposed test does not provide an adequate validation and does not resolve the issue of testing accuracy.

Similarly, another method proposed to validate surface EMG decomposition, the reconstruct-and-test procedure (17, 81), involves generating a synthetic signal from the decomposed trains of action potentials and reapplying the same decomposition method to this synthetic signal to which noise is added with a power similar to that of the residual in the first decomposition (see Fig. 9 in Ref. 16). The accuracy with which each train of action potentials is identified is then estimated as the percentage of action potentials that are identified by the two decompositions with respect to the total. Because validation of the algorithm depends on the initial results and on the residual obtained with the decomposition algorithm, this method does not validate the accuracy of a decomposition algorithm, as discussed in detail by Farina and Enoka (40). The main problem with this method is that it will indicate high accuracy when the residual is minimized, which makes the two decomposed signals very similar, without necessarily maximizing the accuracy of the discrimination. Our concern is not on the decomposition algorithms that have been developed but rather the tests that have been used to establish their validity.

The only current reliable approach to assess the accuracy of a surface EMG decomposition algorithm is a method derived from the two-source approach, previously described by Mambrito and De Luca (71) for decomposing intramuscular EMG signals. This method consists of measuring surface and intramuscular EMG signals concurrently, decomposing the two
signals with two appropriate algorithms, and then determining the rate of agreement in the discharge times discriminated from the two recordings (Fig. 8). This approach does not consider the intramuscular decomposition to be the gold standard but rather is based on the assumption that the result is likely correct (the probability of making the same mistake is low) when two independent decomposition algorithms are applied to two different signals and produce similar results. The accuracy obtained with this approach, which can be expressed as rate of agreement, is not biased to one of the decomposition methods (intramuscular vs. surface) but characterizes the relative performance of the two algorithms. The method developed by Holobar and Zazula (51) to decompose surface EMG signals has been tested with this approach for several muscles and conditions (52, 53, 55), including signals from pathological subjects (54). More recently, the decomposition algorithm proposed by DeLuca et al. (14) has been tested with the same approach for low contraction forces (<15% MVC) (58).

Although more limited than an experimental validation, the use of simulated signals also provides some confidence in the accuracy of a decomposition algorithm. Simulations of surface EMG signals have indeed become so accurate and realistic (5, 32, 19, 70) that their use is a valuable initial step in the process of validating a decomposition method (52).

Conclusions

Use of surface EMG signals to characterize neural activation is complicated by the signal containing information about both the neural drive to muscle and the electrical properties of muscle fibers. Amplitude, spectral, and coherence analysis of the surface EMG are influenced by this and other limitations, as we have detailed in the current review and its predecessor (31). Alternatively, much work over the last decade has expressed increasing confidence in the capacity of some algorithms to decompose a surface EMG signal into its constituent trains of motor unit action potentials. If valid, this approach would permit the direct extraction of information about the neural drive to muscle from surface EMG recordings, with advantages over intramuscular EMG due to the noninvasive approach and the possibility of identifying a greater number of motor units. However, valid accuracy tests involving concurrent intramuscular and surface EMG recordings need to be performed.

ACKNOWLEDGMENTS

We sincerely thank Drs. Francesco Negro and Jakob L. Dideriksen for discussions on topics presented in this review article and for help in generating the data for some of the figures.

GRANTS

We acknowledge financial support by the European Research Council Advanced Grant DEMOVE (Contract No. 267,888; to D. Farina), by Compagnia di San Paolo and Fondazione CRT (to R. Merletti), and by the National Institute of Child Health and Human Development (to R. M. Enoka).

DISCLOSURES

No conflicts of interest, financial or otherwise, are declared by the author(s).

AUTHOR CONTRIBUTIONS


REFERENCES

44. Fuglevand AJ, Winter DA, Patla AE.
43. Farina D, Negro F, Dideriksen JL.
35. Farina D, Cescon C, Negro F, Enoka RM.
33. Farina D.
32. Gazzoni M, Farina D, Merletti R.
Synthesis