Principal Component Analysis of Electromyographic Signals: An Overview

Gianfranco Bosco[*,1,2,3]

[1]*Department of Neuroscience and* [2]*Center of Space Biomedicine, University of Rome Tor Vergata, 00133 Rome, Italy*

[3]*Department of Neuromotor Physiology, Santa Lucia Foundation, 00179 Rome, Italy*

Abstract: Surface electromyography (EMG) is a widely used, straight-forward, technique which allows to investigate patterns of neuromuscular activation. In contrast to the relative simplicity of the recording technique, the analysis of the derived electric signals may be rather sophisticated. The last decade, in particular, has been characterized by the development of a several quantitative approaches to the analysis of the EMG signals. The common principle underlying these analyses is the decomposition of the EMG signal waveforms in a small set of basis waveforms that capture most of the relevant features of the source EMGs and define a low-dimensional space on which the original EMG activation patterns can be represented as vectors. This could be particularly useful when the aim is to classify quantitatively EMG patterns recorded across muscles or from the same muscle across several motor tasks. Within this framework, this article will be focused on one of these approaches, the Principal Component Analysis, which has a strong potential for large scale diffusion both in research and clinical settings because of its conceptual simplicity and high practicality. The intent is to provide an overview/tutorial of the PCA applied to surface EMG signals, first by outlining the main methodological aspects and, then, by drawing examples from the movement control literature where PCA has been used effectively to gain insight on the neural processes that may underlie the control of common actions of our motor repertoire such as arm pointing and gait.

Keywords: Electromyography, Principal Component Analysis, correlation analysis, gait, arm movements.

INTRODUCTION

Surface electromyography (EMG) is a relatively simple technique, widely used in both the clinical and the research field, which allows investigating the patterns of neuro-muscular activation by applying pair of electrodes on the skin surface above the belly of selected muscles. In spite of the overall simplicity of standard surface EMG methods, the analysis of the derived electric signals may be rather complex and sophisticated. One reason for this complexity stems from the inherent nature of the surface EMG signals, which represents the electrical activity of a population of motor units. During motor behavior the population of motor units of a given muscle may be active simultaneously generating an easily recognizable burst-like waveform in the EMG signals derived from the skin surface. In other cases, however, motor units within a muscle, especially in bi-functional muscles, may be activated with different timing and intensity, generating complex temporal waveforms of activation. Moreover, when studying the pattern of activity derived from several muscles throughout the body it becomes critical to find analytical means of comparing and describe synthetically the EMG activity across muscles. Conventional analyses of EMG signals, particularly in the clinical setting, are rather descriptive and generally focused on determining the timing of onsets and peaks as well as the duration of EMG bursts and on characterizing the intensity of muscle activation by defining indexes like the EMG peak amplitude or the integral of the EMG burst. On one hand these analyses do not require much computational power and can be carried out rather easily by a clinician and/or researcher; on the other hand they offer a limited description of the potential complexity of the EMG activation pattern, especially when analyses of EMG waveforms across muscles are to be performed. During the last decade, in parallel with the rapid rise of the computational power available through personal computers, a number of quantitative analyses of the EMG signals have been developed in order to overcome the limitations of the conventional ones [1]. The common principle underlying these analyses is the decomposition of the EMG signal waveforms in a small set of basis waveforms that capture most of the relevant features of the original EMG activation patterns. These analyses are particularly useful when the aim of the study is to classify quantitatively EMG patterns recorded across muscles or from the same muscle across several motor tasks. The reduced number of basis waveforms extracted by means of these analyses define, in fact, a low-dimensional space on which it is possible to examine the distribution of the original EMG activation patterns. The waveform decomposition can be done in the frequency or temporal domain. Examples of waveform decomposition in the frequency domain are represented by Fourier analysis [2] and, in part, by wavelet analysis [3, 4]. Most methods that have been developed for EMG waveform decomposition operate in the temporal domain and are essentially variants of the factor analysis [5]. These include Principal component analysis (PCA), Independent component analysis (ICA), singular value decomposition (SVD) and non-negative matrix factorization

*Address correspondence to this author at the Department of Neuroscience and CSBM, University of Rome "Tor Vergata", Via Montpellier, 1 00133 Rome, Italy; Tel: +390651501483; Fax: +390651501482;
E-mail: bosco@med.uniroma2.it

1874-9437/10

[6-11]. Although there is no ultimate factorization method and the choice among them should be carefully evaluated on the basis of the features inherent to dataset to analyze, PCA stands out both for its conceptual simplicity and also for practical reasons since it is supported by most commercial statistical software packages or it can be programmed rather easily. For these reasons, among other methods, PCA has the strongest potential for a large scale diffusion both in research and clinical settings to analyze large datasets of EMG signals derived from multi-electrode recordings. Therefore, this article is intended to provide an overview of the PCA applied to surface EMG signals, first by outlining the main methodological aspects in the form of a brief tutorial and then by drawing some examples from the movement control literature where PCA has been used effectively to gain insight on the neural processes that may underly the control of common actions of our motor repertoire such as arm pointing and gait.

STATISTICAL BASIS OF PCA

Principal component analysis is a method to extract from a large dataset of waveforms (specifically, the EMG signals) a smaller number of waveforms, the basis waveforms or principal components, that describe the most common features represented in the dataset. In statistical terms, the principal components are the factors explaining most of the dataset variance [5, 12].

Covariance and Correlation

The PCA is based on the concepts of covariance and correlation. The covariance between two variables (X, Y) represents a measure of how much the two variables change together and it can be defined as:

$$\mathbf{cov}(X, Y) = E\{[X - E(X)][Y - E(Y)]\} \quad (1)$$

where E is the mean value.

Correlation is related to covariance (see eq. 2) and it represents a measure of the linear relationship between the two variables that is, how much one variable varies as a function of the other (Morrison, 1990). It can be defined as:

$$\mathbf{cor}(X, Y) = \mathbf{cov}(X, Y)/[\mathbf{sd}(X)\ \mathbf{sd}(Y)] \quad (2)$$

where **sd** is the standard deviation. Note that because of the **sd** terms at the denominator, the correlation, unlike the covariance, is not influenced by the oscillations in amplitude of the two variables but only by their temporal relationship. Correlation between two variables is quantified by the Pearson product-moment correlation coefficient (r), which can assume values between -1 and 1. A positive correlation (or covariance) indicates that the two variables have similar trends while negative correlation (or covariance) signifies that the two variables have specular trends. If the two variables are unrelated the correlation (or covariance) coefficient is equal to zero. Fig. (**1**) illustrates examples of two variables which can be strongly correlated (r = 1), not correlated (r = 0) or anti-correlated (r = -1).

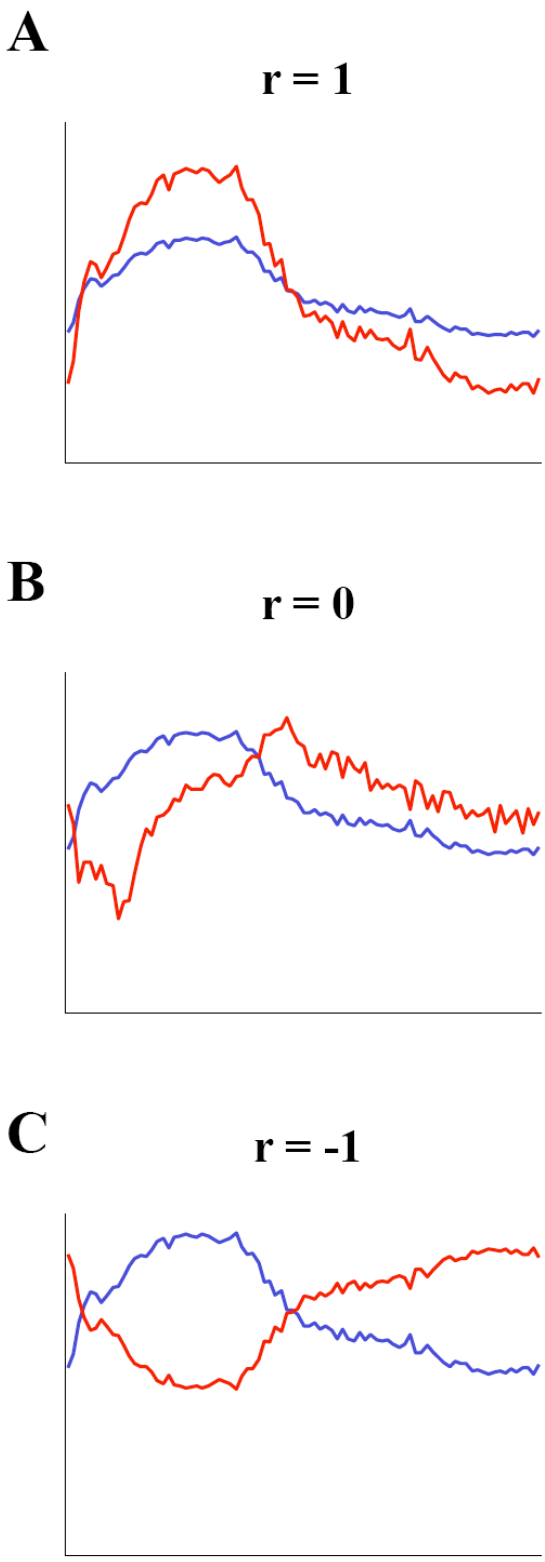

Fig. (1). Correlation between two variables. (**A**) Two variables showing maximal positive correlation (r = 1). (**B**) Uncorrelated time-series (r = 0). (**C**) Anti-correlated time series: the two variables have specular trends (r = -1).

Computing the Principal Components

The first step to perform a PCA is to compute either the covariance or the correlation matrix of the dataset variables, that is the EMG signals. From the definitions of correlation and covariance mentioned above, it follows that PC extracted from a covariance matrix will be dominated by signals with larger amplitudes, whereas those extracted from a correlation matrix will be influenced only by the temporal relationships among the original time-series. In effect, the application of PCA to EMG signals is mostly concerned with identifying common temporal features across EMG signals and therefore a correlation matrix represents the elective choice. In order to compute the correlation (covariance) matrix, the raw EMG signals must be preprocessed by means of rectification, low-pass filtering (cut-off frequency: 10-20 Hz) and normalization of the time-base (for example, if gait data are being processed, EMG signals could be normalized to the fraction of the gait cycle).

In essence, the PCA extracts from the correlation (covariance) matrix:

1. the eigenvectors namely, the directions of the orthogonal axes, which account for most of the dataset variance;
2. the eigenvalues, i.e, the scalar component of the eigenvectors, which indicate the fraction of the total variance accounted for by each eigenvector;
3. the principal components or factor scores (PC), which represent the waveforms associated to each eigenvector/eigenvalue;
4. the weighting coefficients or factor loadings (FL), which represent the Pearson correlation coefficients between the principal components and each original EMG waveform so that the original EMG signal could be reconstructed by the weighted sum of the principal components:

$$EMG = \sum_{1}^{n} pc * fl \qquad (3)$$

The PCs extracted with this method are ordered on the basis of the fraction of variance explained and usually, depending on the complexity and the size of the original dataset, 2 to 5 PCs can account for up to 80-90% of the total variance. This represents a remarkable reduction of the dimensionality of the dataset. In fact, the low-order PCs define a low-dimensional space on which the original dataset of EMG signals could be represented by using the weighting coefficients as vector coordinates. In the PC space, EMG signals that share similar temporal components will tend to cluster together whereas those that show different temporal patterns will be far apart. Fig. (**2**) illustrates one example of such PC space representation. The first two PCs extracted from a dataset of 25 original waveforms accounted for more than 80% of the total variance and the original waveforms have been plotted as vectors using their weighting coefficients in the bi-dimensional space defined by these two PC waveforms. Note that the original waveform shown in the inset is heavily weighted on the first PC.

Rotation of the Principal Components

The PC waveforms extracted using the procedure described above represent a mere statistical description of the common features across the dataset and may not have any particular functional significance. In order to improve the interpretability of the results of a PCA, several algorithms that perform iterative rotations of the PC axes have been developed with the aim of forcing the resulting PC waveforms to be more adherent to the original dataset. PC rotations are subdivided in two families, orthogonal and oblique rotations, depending on whether or not the orthogonality (i.e. the independence) of the PC axes is preserved. The type of rotation that is found to be employed most often with PCA of EMG signals is the Varimax rotation. This is an orthogonal rotation of the PC space

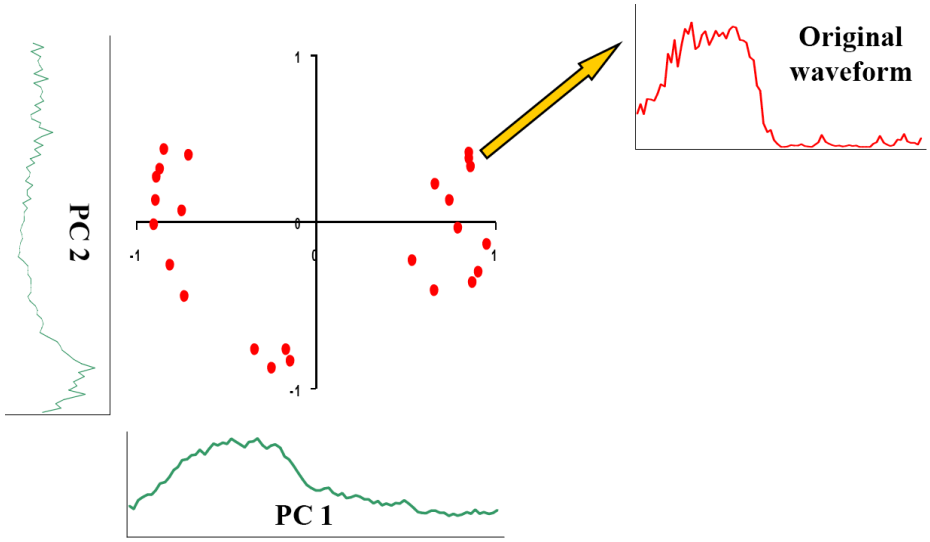

Fig. (2). Principal components space. Principal components were extracted from a dataset of 25 waveforms. The first two PCs explain more than 80% of the total variance and their waveforms are graphed next to the corresponding axes. The original time series are represented as vectors in the bi-dimensional PC space (red circles). Inset shows the original waveform associated with the data-point indicated by the yellow arrow.

which redistributes the variance explained more evenly among the PCs by forcing the weighting coefficients to either maximal (1 or -1) or null values (see Fig. **3**). Consequently, each original EMG signal will be described by fewer PCs and the PC waveforms will be more similar to the original EMG waveforms, making the results of the PCA potentially more interpretable.

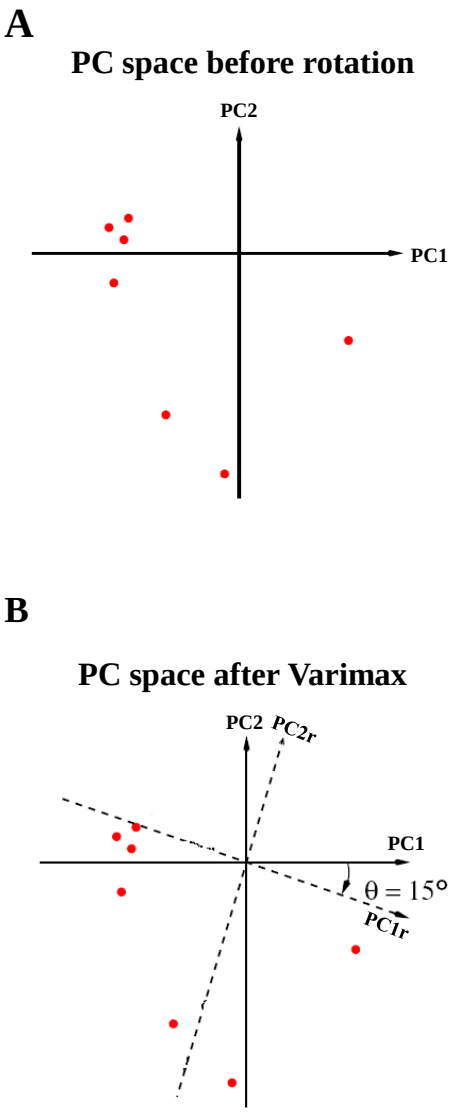

Fig. (3). Varimax rotation. (**A**) PC space before the application of the varimax rotation. (**B**) PC space after varimax rotation. Rotated axes are dashed and the rotation angle is indicated on the right side of the graph. Note that after varimax rotation, data-points representing the original time-series (red circles) lay closer to the PC axes.

WHAT CAN BE LEARNED FROM PCA OF ELECTROMYOGRAPHIC DATA?

In sum, PCA is a powerful technique that identifies common temporal patterns across large datasets of time-series like EMG signals and defines a low-dimensional space on which the original signals could be represented as vectors and classified. Then, how can PCA be used effectively to analyze EMG patterns? There could be at least two potential applications of this method with clear and rather insightful examples coming from the motor control literature. One possible application of the PCA would be to study the electromyographic activity of individual muscles recorded during several variants of one motor action (or even different motor behaviors) and characterize the temporal patterns of activity associated to different components of the motor action. A study by Flanders and Herrmann [6], which pioneered the use of PCA on EMG data, exemplifies this first type of PCA application. These authors recorded the EMG activity of several arm muscles while human subjects were instructed to perform pointing movements at various speeds to targets in different directions. They performed PCA of the EMG signals recorded from individual muscles during movements to one target at different speeds with the aim of identifying components of muscle activity that could be related to movement speed. PCA identified only two components of muscle activity that explained most of the variance in the EMG patterns. By rotating the PC space, Flanders and Herrmann found that the two PCs were represented differently in the EMG activity associated to pointing movements at different speeds. The first PC contributed equally to the EMG activity at different speeds, since its weighting coefficients did not vary with movement time. In contrast, the contribution of the second PC scaled with movement time, as its weighting coefficients showed a clear monotonic trend. This result was interpreted by the authors as evidence for two separate premotor signals driving the motoneuronal pools during arm movements. One speed-independent signal, described by the first PC, would be responsible for generating muscle forces to counteract the effect of gravity, whereas the other premotor signal identified by the second PC would produce phasic muscle activation which scales with movement speed.

Another potential application of PCA relates to the analysis of patterns of electromyographic activity recorded by means of multi-electrode systems (up to 32 electrodes) from many muscles throughout the body during one or more motor actions. PCA, then, may be used to identify spatial-temporal neuromuscular synergies underlying the motor behavior. In this respect, Ivanenko and colleagues have performed a series of insightful studies using PCA with varimax rotation to analyze EMG data recorded from 16 to 32 muscles throughout the body during several locomotion conditions which included walking, running, various levels of body weight support (BWS) and coordination of locomotion with voluntary actions, such as kicking a ball or grasping an object [7, 13, 14, 15]. They identified five basic component waveforms that explained about 90% of the total variance across different muscles activation waveforms during normal gait. These component waveforms tended to be timed relative to the foot lift-off and were invariant with respect to walking speed and body weight unloading. Moreover, by using correlation analyses, it was found that two of the component waveforms were systematically related to the foot kinematics, suggesting that few spinal oscillators can control the limb and the trunk muscles to produce the locomotion kinematics [7, 13]. Interestingly, the five basic muscle activation waveforms were invariably present also when the locomotion motor program had to be coordinated with some additional voluntary action (kicking a ball, stepping over an obstacle or grasping an object from the

ground). The component waveforms, however, were weighted differently across muscles in the different locomotion conditions and a separate activation component, temporally related to the voluntary task, could be also present. According to the authors of this study, these results imply that coordination of locomotion with the voluntary task was achieved with a combination of distinct temporal activation patterns for locomotion and the voluntary task, consistent with the idea that compound movements may be accomplished by virtue of a superposition of motor programs [14]. Finally, PCA was also used to analyze EMG data recorded from spinal patients supported with a BWS system and trained to step on a treadmill. It was found that, at the end of the training program, patients produced similar foot kinematics to that of control healthy subjects but using different activations of individual muscles [13]. However, PCA indicated that in patients the basic set of five temporal components was preserved and that the large variability of muscle patterns observed between controls and patients could be explained by flexible combinations of the same basic component waveforms. This result suggested further that the basis waveforms identified by the PCA could be related to control signals output by spinal pattern generators.

In conclusion, the applications of PCA of EMG data outlined here highlight the potential for this method (as well as for similar waveform decomposition methods) to capture features from surface EMG signals that can provide insight not only on the activation state of motoneurons, but also on the nature of the premotor control signals, opening new windows for both neurophysiological and clinical/ rehabilitation studies.

ABBREVIATIONS

BWS = Body Weight Support

EMG = Electromyography

FL = Factor Loadings

ICA = Independent Component Analysis

PC = Principal Components

PCA = Principal Component Analysis

SVD = Singular Value Decomposition

REFERENCES

[1] von Tscharner V, Herzog W. EMG. In: Nigg B, Herzog W, Eds. Biomechanics of the Musculo-Skeletal System, 3rd Edition. Hoboken, NJ (USA): John Wiley & Sons 2007.

[2] Rainoldi A. Spectral properties of the surface EMG can characterize/do not provide information about motor unit recruitment strategies and muscle fiber type. J Appl Physiol 2008; 105(5): 1678.

[3] Kilby J, Gholam Hosseini H. Wavelet analysis of surface electromyography signals. Conf Proc IEEE Eng Med Biol Soc 2004; 1: 384-7.

[4] Maitrot A, Lucas MF, Doncarli C, Farina D. Signal-dependent wavelets for electromyogram classification. Med Biol Eng Comput 2005; 43(4): 487-92.

[5] Gorsuch RL. Factor Analysis. Hillsdale, NJ (USA): Erlbaum 1983.

[6] Flanders M, Herrmann U. Two components of muscle activation: scaling with the speed of arm movement. J Neurophysiol 1992; 67(4): 931-43.

[7] Ivanenko YP, Poppele RE, Lacquaniti F. Five basic muscle activation patterns account for muscle activity during human locomotion. J Physiol 2004; 556(Pt 1): 267-82.

[8] Cheung VC, d'Avella A, Tresch MC, Bizzi E. Central and sensory contributions to the activation and organization of muscle synergies during natural motor behaviors. J Neurosci 2005; 25(27): 6419-34.

[9] Tresch MC, Cheung VC, d'Avella A. Matrix factorization algorithms for the identification of muscle synergies: evaluation on simulated and experimental data sets. J Neurophysiol 2006; 95(4): 2199-212.

[10] Naik GR, Kumar DK, Arjunan SP, Palaniswami M, Begg R. Limitations and applications of ICA for surface electromyogram. Conf Proc IEEE Eng Med Biol Soc 2006; 1: 5739-42.

[11] Tabernig CB, Acevedo RC. M-wave elimination from surface electromyogram of electrically stimulated muscles using singular value decomposition: preliminary results. Med Eng Phys 2008; 30(6): 800-3

[12] Morrison DF. Multivariate Statistical Methods. New York (USA): McGraw-Hill 1990.

[13] Ivanenko YP, Grasso R, Zago M, *et al*. Temporal components of the motor patterns expressed by the human spinal cord reflect foot kinematics. J Neurophysiol 2003; 90(5): 3555-65.

[14] Ivanenko YP, Cappellini G, Dominici N, Poppele RE, Lacquaniti F. Coordination of locomotion with voluntary movements in humans. J Neurosci 2005; 25(31): 7238-53

[15] Cappellini G, Ivanenko YP, Poppele RE, Lacquaniti F. Motor patterns in human walking and running. J Neurophysiol 2006; 95(6): 3426-37.

Received: August 28, 2009 Revised: March 21, 2010 Accepted: March 21, 2010