Identification of regional activation by factorization of high-density surface EMG signals

Gallina, Alessio; Garland, S. Jayne; Wakeling, James M.

DOI:
10.1016/j.jelekin.2018.05.002

Title: **Identification of regional activation by factorization of high-density surface EMG signals: a comparison of Principal Component Analysis and Non-negative Matrix factorization**

Alessio Gallina[1], S Jayne Garland[2,3], James M. Wakeling[4]

[1] Graduate Programs in Rehabilitation Sciences, University of British Columbia, Vancouver, V6T 1Z3

[2] Department of Physical Therapy, University of British Columbia, Vancouver, V6T 1Z3

[3] Faculty of Health Sciences, University of Western Ontario, London, N6A 5B9

[4] Department of Biomedical Physiology and Kinesiology, Simon Fraser University, Canada, Burnaby, V5A 1S6

Corresponding author: S. Jayne Garland, PhD PT

University of Western Ontario, Faculty of Health Sciences

200 Arthur & Sonia Labatt Health Sciences Bldg, 1151 Richmond St

London, ON Canada N6A 5B9

Email: jgarland@uwo.ca

Word count: 4436

Figures: 4

Table: 1

All authors have seen and approved the final version of the manuscript being submitted. They warrant that the article is the authors' original work, hasn't received prior publication and isn't under consideration for publication elsewhere.

The authors declare no conflicts of interest.

ABSTRACT:

In this study, we investigated whether principal component analysis (PCA) and non-negative matrix factorization (NMF) perform similarly for the identification of regional activation within the human vastus medialis. EMG signals from 64 locations over the VM were collected from twelve participants while performing a low-force isometric knee extension. The envelope of the EMG signal of each channel was calculated by low-pass filtering (8 Hz) the monopolar EMG signal after rectification. The data matrix was factorized using PCA and NMF, and up to 5 factors were considered for each algorithm. Association between explained variance, spatial weights and temporal scores between the two algorithms were compared using Pearson correlation. For both PCA and NMF, a single factor explained approximately 70% of the variance of the signal, while two and three factors explained just over 85% or 90%. The variance explained by PCA and NMF was highly comparable ($R > 0.99$). Spatial weights and temporal scores extracted with non-negative reconstruction of PCA and NMF were highly associated (all $p < 0.001$, mean $R > 0.99$). Regional VM activation can be identified using high-density surface EMG and factorization algorithms. Regional activation explains up to 30% of the variance of the signal, as identified through both PCA and NMF.

KEYWORDS: EMG; factorization; regionalization; quadriceps; vastus; neuromuscular control.

INTRODUCTION:

Variations in the orientation, insertion and architecture of muscle fibres can be observed within most human muscles. The human vastus medialis (VM), for instance, inserts on different regions of the common knee extensor tendon and along the medial edge of the patella (Holt et al., 2008; Peeler et al., 2005), and its fibers are more obliquely oriented in the distal than in the proximal region of the muscle (Gallina and Vieira, 2015; Peeler et al., 2005; Smith et al., 2009). These anatomical variations may render different regions within the VM mechanically more effective at producing force along different directions, i.e.: knee extension for the proximal fibers, patellar translation and rotation for the distal ones (Lin et al., 2004; Wilson and Sheehan, 2010). As motoneurons recruited at low force levels tend to innervate muscle units occupying small territories within the VM (Gallina and Vieira, 2015; Gootzen et al., 1992), it may be possible for motor units located in different VM regions to receive an uneven distribution of neural drive (Tenan et al., 2016, 2013) or afferent feedback (Gallina et al., 2017), resulting in regional activation such as that observed in dynamic contractions (Gallina et al., 2016).

Surface electromyography is a technique commonly used to describe the timing and intensity of muscle activation. While myoelectric activity sampled with a single pair of electrodes is generally considered to represent the activation of the whole muscle, regional activation may lead to less representative EMG estimates than previously thought (Gallina et al., 2011). Vieira and colleagues (Vieira et al., 2015) showed a heterogeneous increase in sEMG amplitude across the human gastrocnemius in response to incremental electrical stimulation, demonstrating how detection of sEMG from a single location within a muscle may offer limited information on the intensity of the whole-muscle activation. The EMG envelope is widely used as an estimate of timing and amplitude of muscle activation in dynamic, isometric (Negro et al., 2009) and quasi-isometric tasks (Gallina et al., 2016a;

Masani et al., 2003); however, the effects of regional muscle activation on this EMG parameter, if any, are poorly understood.

Factorization algorithms are used to extract common patterns of EMG activity across several muscles during functional tasks (i.e., muscle synergies, (Cheung et al., 2005; Tresch et al., 2006)). Similarly, when applied to recordings performed using a grid of closely spaced surface EMG electrodes (High-Density surface EMG, HDsEMG), algorithms such as principal component analysis (PCA; Staudenmann et al., 2013a, 2009) and Non-negative Matrix Factorization (NMF; Gazzoni et al., 2014; Huang et al., 2015; Muceli et al., 2013) were shown to be able to factorize the HDsEMG signals in clusters of electrodes that have similar profiles of temporal activation. These algorithms may be useful to describe how well the original signal can be reconstructed, assuming absence of regional activation (a single factor that fluctuates similarly across channels) as opposed to more complex spatiotemporal patterns (a larger number of factors). Additionally, factorization algorithms may provide objective information on the spatiotemporal characteristics of the activation of regions within the VM. A direct comparison of the factors obtained using PCA and NMF will provide information on whether spatiotemporal patterns of regional activation are identified similarly by the two algorithms, and whether one algorithm is preferable for the identification of specific neuromuscular activation features.

The purpose of this study was to investigate whether PCA and NMF provide similar results when used to identify regional muscle activation from HDsEMG signals. We hypothesized that the identification of multiple regions within the VM activation patterns would be a robust physiological feature that was independent of the analysis method used (NMF or PCA). We also hypothesized that a single factor will explain only part of the total variance of the signal, indicating that a single EMG envelope estimate may not be representative of the whole VM muscle.

METHODS:

Participants

Twelve healthy people (4 F; 33 ± 13 years old; height: 179 ± 9 cm; weight: 75 ± 12 kg) participated in the study. Participants did not report any pain or neuromuscular disorders at the time of the data collection. The experimental protocol was approved by the UBC research ethics board, and each participant signed a consent form before starting the experimental session.

Protocol

The position and orientation of the EMG grid was defined on the basis of participant-specific anatomy of their vastus VM. Participants sat on an adjustable chair (Biodex Medical, Shirley, NY, USA). Their right leg was examined with an ultrasound scanner (Telemed, Vilnius, Lithuania) to define the mediolateral edges of their VM muscle. The position of the innervation zone across the VM, which can be observed in surface EMG signals as phase opposition of propagating action potentials (Gallina et al., 2013), was identified using a dry array (16 electrodes spaced 10 mm; OTBioelettronica, Torino, Italy) during a manually-resisted low-force knee extension. After cleaning the skin with a small amount of abrasive paste, a 64-channel grid (13x5 electrodes, one missing corner, 8 mm interelectrode distance, OTBioelettronica, Torino, Italy) was placed on the VM according to the identified anatomical landmarks; the grid was centered between the medial and lateral edges of the VM, and oriented to have VM innervation zone aligned between the second and third columns of electrodes (fig. 1). The grid was placed on the skin and kept in place with double-adhesive foam; conductive paste was used to ensure a good reading of the electrodes. Two adhesive electrodes (1 cm in diameter), placed on the patella and on the femoral medial epicondyle, were used as reference electrodes for the surface EMG system. The task consisted of a low-force isometric knee extension, performed with hip and knee joint flexion angles

of 90 degrees. Participants were provided with visual feedback of the EMG activation and were asked to reach a target amplitude of 50 μV (average of the 5 highest amplitude value of the electrode grid) in approximately 10 seconds, and to maintain that muscle activation for 30 s. After the protocol, the EMG activation during two maximal voluntary contractions was recorded; this showed that an activation of 50 μV was approximately 5-10% of the maximal activation. Electromyographic signals were collected in monopolar modality at 2048 Hz using an amplifier (128-channel EMG-USB; OTBioelettronica, Torino, Italy). As thickness of the tissue interposed between the electrodes and the muscle is a factor known to influence EMG estimates, ultrasound was used to collect one image from the proximal and distal regions of the VM (approximately rows 3 and 10).

Data Analysis

The distance between skin surface and the most superficial region of the VM was measured from the ultrasound images using ImageJ (National Institutes of Health, Bethesda, Maryland, USA). After band-pass filtering (10-400 Hz), a 30s epoch of the EMG signal was used for the analysis. The main analysis used EMG envelopes calculated by full-wave rectifying and then low-pass filtering (Butterworth, 4^{th} order, 8 Hz) the EMG signal from each channel of the grid (fig.1); EMG envelopes were then down-sampled to 32 Hz, resulting in 1920 time points. Factorization algorithms were run on a matrix **M** of 64 (channels) by 1920 (time samples) for each participant separately. As the number of components may depend on the cut-off frequency used to calculate the envelope of the EMG signal (Hug et al., 2012), a secondary analysis compared signals filtered at different frequencies (2, 4, 6, 8, 12, 20 and 50 Hz), similarly to those used in other studies (Cronin et al., 2015; Gallina et al., 2016b; Huang et al., 2016; Masani et al., 2003; Staudenmann et al., 2013b; Vieira et al., 2010b).

Non-negative Matrix Factorization (Lee and Seung, 1999) was run using a code from the Statistics Toolbox of Matlab 2014. In brief, the envelope matrix was factorized in N factors, each consisting of N

spatial weights and 1920 temporal scores. The algorithm performs a low-rank approximation of the matrix **M** in spatial weights $\mathbf{SS_{NMF}}$ and temporal scores $\mathbf{TC_{NMF}}$. The factorization can be described as:

$$M = SS_{NMF} * TC_{NMF} + \varepsilon$$

where * denotes a matrix product and **ε** is the error between the original matrix **M** and the factorized matrix $\mathbf{SS_{NMF}} * \mathbf{TC_{NMF}}$. Through an iterative process, the NMF algorithm reduces **ε** so that $\mathbf{M \approx SS_{NMF} * TC_{NMF}}$. As the algorithm starts from random initial states of $\mathbf{SS_{NMF}}$ and $\mathbf{TC_{NMF}}$, the processing was repeated 50 times starting with different random states each time. The factorization that resulted in the lowest error was chosen. As the number of factors has to be set prior to evaluating the NMF, the algorithm was run with a number of factors between 1 and 5.

Principal component analysis (Joliffe 1986) was run using a custom-made computer program. In brief, the EMG envelope matrix **M** was factored into 64 principal components (PCs), each consisting of 64 weights and 1920 coefficients. Weights were calculated as the eigenvectors **ζ** of the covariance matrix of **M**. Coefficients were calculated as $\boldsymbol{\zeta}^{T} * \mathbf{M}$, which is the matrix product between the transposed eigenvectors and the EMG envelope matrix. Principal components were sorted according to their eigenvalues. As the information of the spatial weights or temporal scores may be equivalent between the two factorization techniques, but expressed differently because of the non-negative constraint imposed by NMF and not by PCA, non-negative spatial weights and temporal scores were reconstructed from the PCA, in a similar manner to a previous study (Hodson-Tole and Wakeling, 2007). This is indicated in the text as reconstructed PCA (PCAr). In brief, the spatial weights of PC1r and PC2r were calculated as:

$$PC1wR = PC1w + C\ PC2w$$

where PC1w and PC2w are the PCA weights, PC1wR is the weights of the reconstructed (non-negative) PC1, C was identified using an iterative process and corresponds to the largest value that results in PC1wR having all positive values. PC2wR was calculated with the same formula, but C was defined as the smallest negative value that results in PC2wR having all positive values. The temporal coefficients were identified using an iterative process based on least square difference. For each time sample, the error was calculated as:

$$error == \sqrt{\sum (ENV - D1\ PC1wR + D2\ PC2wR)^2}$$

Where ENV is the original envelope for each time point (a matrix of 13x5 values), PC1wR and PC2wR are the reconstructed (non-negative) weights calculated from the PCA, D1 and D2 were values identified using an iterative process that minimized *error* at each time point.

These factors, called spatial weights and temporal scores, were used instead of the original PCA factors when direct comparisons between spatial and temporal factors extracted with the two factorization techniques were made.

For each number of factors, the quality of the reconstruction of the signal was assessed by comparing the variance of the signal reconstructed to that of the original signal. For both factorization techniques, the variance of the reconstructed signal was calculated using the coefficient of determination, calculated as:

$$CD = 1 - \frac{SSR}{SST}$$

$$SSR = \sum_{i=1}^{n} \left(y_i - f(x_i)\right)^2$$

$$SST = \sum_{i=1}^{n} (y_i - \bar{y})^2$$

where CD is the coefficient of determination, SSR is the sum of squared residuals, SST total sum of variance of all channels of the original signal **M**. The average variance explained by 1-5 factors was calculated for both PCA and NMF.

Comparison of factors extracted with PCA and NMF

The similarity between factors extracted with NMF and PCA was assessed by correlating the variance explained, the spatial weights and temporal scores. For each number of factors, concurrent validity between NMF and PCA was determined by calculating the association between the variance explained by the two techniques. As a secondary analysis, this was repeated for the EMG envelopes created by filtering the EMG signals at different frequencies. For NMF run with N = 2, spatial weights and temporal scores extracted with NMF and PCAr were compared according to their position within the VM (e.g.: spatial weights of the proximal NMF factor were correlated to spatial weights of the proximal PCAr factor). Each factor was determined to be encoding information from the proximal or distal VM based on the location of its active area (Gallina et al., 2016), which was defined as the "Rows" coordinate of the barycenter of the channels higher than 70% of the maximum spatial weight (Vieira et al., 2010a; fig.2). The barycenter was calculated as:

$$\frac{\sum Wch\ POSch}{\sum Wch}$$

where *ch* is each channel of the electrode grid, *W* is the value of the spatial weight, *POS* is the Y (proximal-distal) coordinate of the channel.

Comparison of proximal vs. distal NMF factors

As factorization performed with NMF or PCA/PCAr was basically equivalent, this analysis was run on NMF only. For NMF run with N = 2, temporal and spatial independence of proximal vs. distal factors were tested by correlating temporal scores and spatial weights. To identify differences in size of the active area between factors, the percentage of channels with value higher than 70% of the maximum of each spatial weight (Vieira et al., 2010a; fig.2) was calculated and compared between proximal and distal factors to identify the difference in size of the active area within the muscle. The median frequency of the temporal scores was calculated using a Fourier-transform and compared to identify temporal differences associated with regional muscle activation patterns.

Statistics

Differences in the thickness of tissues interposed between skin and muscle between proximal and distal regions of the VM were tested using paired Wilcoxon test. The variance explained for the different number of factors is presented. For the EMG analyses, the Pearson correlation coefficient *R* was used to test association, and paired Student t-tests were used for between-algorithm and proximal-distal comparisons; both tests were run on N = 12 participants. Analyses on the variance explained considered 5 factors for both PC and NMF. As 2 factors were shown to explain most of the variance of the signal (> 85%), further analyses of spatial and temporal features of the factors were run on the first two 2 factors extracted with both PCA and NMF. This choice was also supported by the fact that anatomical (Smith et al., 2009; Peeler et al., 2005), biomechanical (Lin et al., 2004; Wilson and Sheehan, 2010) and motor control studies (Cabral et al., 2017; Rainoldi et al., 2008; Tenan et al., 2016) usually consider the VM as consisting of two regions, proximal and distal. To ensure that the proximal and distal factors explained most of the variance, NMF was run again considering exclusively the channels identified to be the active

area of either factor. The variance explained was calculated as described earlier and reported. The statistical significance was set at $p < 0.05$ or lower according to the Bonferroni correction required.

RESULTS:

The thickness of tissues interposed between skin and muscle was larger ($p < 0.05$) in the proximal (median: 5.9 mm; 25^{th}-75^{th} percentiles: 4.4-7.5 mm) than in the distal VM (median: 4.2 mm; 25^{th}-75^{th} percentiles: 3.5-6.3 mm).

Spatial weights and temporal scores extracted with PCA and NMF

The spatial weights of factors extracted with both algorithms exhibited clear regional organization (fig.1). PC1 always consisted of only positive weights and is an approximation of the mean signal. PC2 described proximal/distal regional activation, having positive spatial weights in the proximal regions and negative spatial weights in the distal region, or vice versa. PC3 and above generally resulted in further regionalization along the longer dimension of the electrode grid, increasing the number of peaks with positive and negative spatial weights. Similarly, a single NMF factor had spatial weights distributed across all channels, whereas running the algorithm with a higher number of factors resulted in clusters of channels progressively more localized along the longer dimension.

Variance explained, comparison of 1-5 factors

The variance explained by different numbers of factors extracted with NMF and PCA was calculated to describe whether a single factor can adequately represent the envelope fluctuations of the VM. As depicted in figure 3, a single factor extracted with either PCA or NMF explained approximately 70% of

the variance of the signal (PCA: 70.9 ± 11.2%; NMF: 70.9 ± 11.2%); two and three factors explained just over 85% (PCA: 87.3 ± 5.6%; NMF: 87.3 ± 5.6%) or 90% (PCA: 92.7 ± 3.9%; NMF: 92.4 ± 3.8%) of the variance, respectively. When the envelopes were obtained by filtering the EMG signal at different frequencies, higher low-pass frequencies generally resulted in lower average variance explained (descriptive analysis; Table 1). Regardless of the filtering frequency, a single component accounted for less than 77% of the variance of the signal (PC1 2-50 Hz: 77-64%; PC2: 88-83%; PC3: 94-89%).

Comparison of PCA and NMF

Factorization with PCA (or PCAr) and NMF was compared by correlating 1) the variance explained for each number of components, 2) the spatial weights and 3) the temporal scores between the two algorithms. When matched for the number of factors, the variance explained by factorizing the EMG envelopes with PCA and NMF was highly comparable ($R > 0.99$ for any number of factors). The variance explained by PCA was significantly larger than that explained by NMF (paired T-tests, all $p < 0.001$; significant if $p < 0.01$ because of Bonferroni correction), although the absolute difference was minimal (1 PC: < 0.001%; 5 PCs < 2%). Spatial weights and temporal scores extracted with PCAr and NMF were highly correlated, for both proximal and distal factors (all $p < 0.001$, mean $R > 0.99$). An example of the EMG envelopes from channels placed proximally to distally in the VM and factors obtained with PCA, NMF and PCAr is shown in figure 2.

Comparison of proximal and distal NMF factors

When NMF was repeated including only the channels identified as active area, separately for the proximal and distal factors, the variance explained was 88.2±4.9%, and this was greater than the variance when a similar number of channels was selected that was evenly distributed across the entire muscle (68%). The location of the centre of the active area of proximal and distal spatial weights for all

participants is shown in figure 4A. Each factor was always represented by a cluster of channels located in a single region of the grid. To investigate temporal and spatial independence of factors extracted with NMF, proximal spatial weights and temporal scores were correlated with distal spatial weights and temporal scores. Spatial weights of proximal vs. distal NMF factors were strongly, negatively associated (mean: R = -0.86 ± 0.11), meaning that the spatial weights reflected fluctuations of components located in different regions of the grid (fig. 2). Temporal scores of proximal vs. distal NMF factors were associated (7/12 positively, 5/12 negatively; $p < 0.001$ in 11/12 participants), but the R^2 was on average small (0.05 ± 0.07; 0.06 ± 0.07 when tested with cross-correlation).

To investigate location-dependent differences in the factors, the size of the active area of the spatial weights and the frequency of the fluctuation of the temporal scores were compared between proximal and distal factors. The active area (channels with spatial weights higher than 70% of the highest spatial weight) was larger for factors representative of proximal than distal VM regions in 9 participants out of 12 (mean: 25.4 ± 4.1% vs. 19.1 ± 6.0%, $p = 0.04$; fig. 4B). The activity within the proximal region fluctuated more slowly (as quantified by the temporal scores) than the activity in the distal region (median frequency: 4.2 ± 0.3 vs. 4.5 ± 0.5 Hz, $p = 0.03$; Fig.4B); this trend was observed in all participants but one.

DISCUSSION:

Regional activation within the vastus medialis in low-force, isometric contractions was identified in this study using high-density surface electromyography and factorization algorithms. Factorization with (non-negative) PCA and NMF resulted in factors with highly comparable amounts of variance explained, as well as spatial and temporal features. As factorization with a single factor explained on

276 average 70% of the variance across channels, a single EMG envelope estimate does not provide
277 complete information on the amplitude fluctuations of the whole vastus medialis EMG during an
278 isometric task. This study also demonstrates that activation is modulated regionally within the VM.

279 A single factor from either the NMF or PCA explained on average 70% of the variance across
280 channels using either factorization algorithm, indicating that more than one factor is necessary to
281 reconstruct an acceptable amount of the variance of the original signal. Up to 30% of the variance was
282 explained by regional variations in muscle activity rather than common fluctuations across the muscle,
283 suggesting that a single bipolar electrode would not fully capture the EMG across the whole muscle.
284 However, it should be noted that a part of this variability may be due to noise. When NMF was applied
285 exclusively to the channels identified as active area for proximal and distal components separately, the
286 variance explained by a single factor was close to 90%, indicating that EMG amplitude fluctuations were
287 similar across channels. While this supports the use of two components instead of one when analyzing
288 EMG signals from the VM, the remaining 10% of unexplained variance could be due to noise or further
289 regionalization within the proximal and distal factors. Indeed, as a larger number of components
290 resulted in factors describing the activation of regions progressively smaller within the VM, it is
291 suggested that the regionalization within the VM is not discrete (e.g.: proximal and distal), but instead
292 continuous. This is in line with studies that showed highly localized stretch reflexes in VM regions as
293 close as 10mm (Gallina et al., 2017), as well as investigations in other muscles (Herrmann and Flanders,
294 1998). This finding suggests that the number of components considered can be varied depending on the
295 spatial resolution needed for the analysis. A lower number of factors can describe the spatiotemporal
296 activation of larger regions within the muscle, while a higher number will provide detailed information
297 on the activation of more localized regions. For instance, if the aim of the analysis is to compare
298 activation of proximal and distal VM, which apply forces to the patella in different directions (Lin et al.,
299 2004), it may be beneficial to factorize the signal using 2 components rather than having activation

profiles for a large number of smaller regions. As expected, envelopes calculated by filtering the EMG signals with lower cutoff frequencies resulted in larger amounts of variance. However, the observation that more than one factor is necessary to reconstruct an acceptable amount of the variance of the original signal holds for any of the frequencies tested; indeed, two components were needed to reconstruct an average value close to 90% for frequencies 2 to 20 Hz (89.7-86.2%), whereas three components were needed for envelopes calculated with a cut-off frequency of 50 Hz (89.3%). In this study, large individual differences in the variance explained by each number of factors were observed, with values ranging between 50% and 85% variance explained for a single component. This is likely due to differences in motor strategies as well as muscle architecture in the heterogeneous sample of participants recruited for this study. It has been previously discussed that the use of HDsEMG and factorization algorithms may help identify between-subject differences in muscle architecture or motor strategies on surface EMG variables (Huang et al., 2015); in line with this view, we report large inter-individual differences in variance explained likely related to different contribution of regional VM activation across participants.

As factors extracted with NMF or (non-negative, reconstructed) PCA resulted in almost equivalent variance explained and spatiotemporal features, it is suggested that these factors reflect information extracted from the signals rather than mathematical artefacts. This concurrent validity is established because the factors are identified in the two analyses based on different processes, i.e.: iterative reduction of residuals between original and reconstructed signal (NMF) and eigenvector decomposition of data covariance (PCA). Although a single factor explained a similar amount of variance between NMF and PCA, the PCA yielded significantly better reconstruction of the original signal; however, this difference was marginal (< 2%), suggesting a largely comparable performance of the two algorithms. Similarly, a recent study (Lambert-Shirzad and Van der Loos, 2017) showed that the number of synergies necessary to reconstruct muscle activation signals during upper limb reaching was similar

regardless of the algorithm used (PCA or NMF). Conceptually, the factorization of HDsEMG signals with NMF can be interpreted as the activation in time of a number of regions within the VM. Instead, PCA outputs a general drive to the muscle (PC1) and regional facilitation/inhibition (PC2, Figure 2). However, the two algorithms result in similar variance explained and similar spatial weights and temporal scores if a non-negative constraint is applied to the PCA reconstructions, suggesting that the two algorithms are able to identify the same patterns of regional activation. Future studies should investigate whether procedures to apply non-negative constrains to PCs different from the one used in this study yield different results. In addition, whether these two algorithms result in similar factorization of other surface EMG dataset is unknown; the similarity between the performance of the two algorithms observed in this study may be due to the characteristics of the dataset, which comprised 64 EMG channels with highly redundant information as many channels were positioned along the same muscle fibers (Staudenmann et al., 2013a).

The strong negative association between spatial weights of proximal and distal factors indicates that the two factors are located in separate spatial locations within the VM, with little overlap (see Figures 2, 4). The spatial information of these factors is consistent with previous studies on the VM. The fact that proximal and distal factors are represented by channels clustered in a muscle region, rather than scattered across the electrode grid, indicates groups of fibers residing in a specific muscle region (Gallina and Vieira, 2015) are more likely to fluctuate their activation together than motor units located far apart. These identified regions are indicative of activations of motor units located more proximally and more distally within the muscle as shown with selective, intramuscular stimulation (Gallina et al., 2016). Finally, visual observation of the spatial weights of each factor have a single peak and a gradual decline in both directions of the grid, similarly to what observed in monopolar recordings over the innervation zone (Kleine et al. 2000; Rodriguez-Falces et al. 2013; Gallina et al. 2016a). Based on the similarities between the spatial representation of factors identified in this study and the anatomical

information from previous studies, it is reasonable to suggest that each factor represents the activation of a group of fibers residing in a region of the VM.

The minimal variance ($R^2 < 0.07$ both for Pearson correlation and cross-correlation) shared by the temporal scores indicates almost independent temporal profiles of the factors identified. Contrary to PCA, NMF does not impose constraints on the temporal independence of the factors; hence, the low variance ($R^2 < 0.07$) observed in this study suggests that fluctuations in the proximal and distal VM are mostly independent in time. However, while the R values were low, the correlation was positively or negatively significantly associated in 11/12 participants, possibly indicating subject-specific motor strategies. As the two factors were centered on the innervation zone of two different VM regions and not along the muscle fiber orientation (Gallina et al., 2016), differences in temporal profiles cannot be due to action potential propagation. Physiologically, these results point to a partially inhomogeneous neural drive to the VM muscle during isometric, low-force contractions. Several mechanisms may contribute to regional differences in neural drive. In the biceps brachii, short-term synchronization of motor units was reported to be higher for motor units located within the same regions than in different compartments of the muscle (Barry et al., 2009). Similarly, recent evidence suggests that common drive may be higher for motor units identified in the same than in different VM regions (Cabral et al., 2017). Regional variations in net neural drive may also stem from uneven reflex inputs or gains; as highly localized stretch reflexes can be observed when VM fibers as close as 10mm are stretched in humans (Gallina et al., 2017), localized 1A afferent feedback may contribute to uneven distribution to synaptic input to VM motoneurons. Through factorization of sEMG amplitude fluctuation, this study supports the fact that motor units located in the VM (and possibly VL) share a common synaptic input (Laine et al., 2015), but at least a part of this neural drive is specific for motor units localized in different muscle regions (Tenan et al., 2016, 2013).

Spatial and temporal features of proximal vs. distal factors were compared to identify region-specific activation patterns within the VM. The proximal VM region was larger and the activation from this region fluctuated more slowly than the distal one, possibly indicating regional differences in anatomy and motor controls although more research is needed. The size of a factor may be related to the number of muscle fibers showing similar fluctuation of activation in time. Although speculative at this point, possible reasons for these differences are: 1) larger number of motor units showing common fluctuations in the proximal than in the distal VM; 2) larger territory size of motor units located proximally than distally; 3) regional differences in thickness of the tissues between recording electrodes and muscle; specifically, the distribution on the skin of surface EMG signals is known to depend on the depth of active motor units (Gallina and Vieira, 2015; Rodriguez-Falces et al., 2013; Roeleveld et al., 1997); for this reason, thicker subcutaneous tissues in the proximal VM may have contributed to more "spread" in the surface EMG, and hence a larger proximal factor. Regional differences in the size of the factors may be due to one or more of the factors mentioned above, or other reasons; from the current data, only speculations can be made. Proximal-distal differences were also observed in the frequency of the temporal profile, the proximal factor fluctuating slower than the distal one, suggesting again regional differences in motor control strategies and/or anatomy; possible contributing factors may be region-specific motor unit synchronization, regional differences in the properties of individual motor unit action potentials, or differences in the frequency of the drive to VM regions possibly associated to the larger proportion of type I motor units in the proximal than in the distal VM (Travnik et al., 2013). These possibilities are, however, speculative at this point.

CONCLUSIONS:

PCA and NMF perform similarly in the identification of regional activation. In low-force isometric contractions of the vastus medialis, a single estimate of EMG envelope explains on average 70% of the

394 variance of the signal collected over several regions across the muscle. Factorization algorithms may be
395 useful to extract common patterns or region-specific activation from HDsEMG recordings, identifying
396 regions whose activation fluctuates largely independently.

397

AUTHOR CONTRIBUTION:

Study design: AG, SJG, JMW. Data collection: AG. Data analysis: AG, JMW; Manuscript preparation: AG, SJG, JMW.

GRANTS:

Alessio Gallina was supported by a Vanier Graduate Canada Scholarship. This study was supported in part by grants from the Natural Sciences and Engineering Research Council of Canada (SJG).

DISCLOSURES:

The authors declare no conflicts of interest.

REFERENCES:

Barry, B.K., Pascoe, M.A., Riek, S., Carson, R.G., Enoka, R.M., 2009. Common input to different regions of biceps brachii long head. Exp. brain Res. 193, 351–9.

Cabral, H. V., de Souza, L.M.L., Mello, R.G.T., Gallina, A., de Oliveira, L.F., Vieira, T.M., 2017. Is the firing rate of motor units in different vastus medialis regions modulated similarly during isometric contractions? Muscle and Nerve 57, 279-286.

Cheung, V.C.K., d'Avella, A., Tresch, M.C., Bizzi, E., 2005. Central and sensory contributions to the activation and organization of muscle synergies during natural motor behaviors. J. Neurosci. 25, 6419–34.

Cronin, N.J., Kumpulainen, S., Joutjärvi, T., Finni, T., Piitulainen, H., 2015. Spatial variability of muscle activity during human walking: The effects of different EMG normalization approaches. Neuroscience 300, 19–28.

Gallina, A., Blouin, J.-S., Ivanova, T.D., Garland, S.J., 2017. Regionalization of the stretch reflex in the human vastus medialis. J. Physiol. 595, 4991-5001.

Gallina, A., Ivanova, T.D., Garland, S.J., 2016. Regional activation within the vastus medialis in stimulated and voluntary contractions. J Appl Physiol 121, 466–474.

Gallina, A., Merletti, R., Gazzoni, M., 2013. Innervation zone of the vastus medialis muscle: position and effect on surface EMG variables. Physiol. Meas. 34, 1411–22.

Gallina, A., Merletti, R., Vieira, T.M.M., 2011. Are the myoelectric manifestations of fatigue distributed regionally in the human medial gastrocnemius muscle? J. Electromyogr. Kinesiol. 21, 929–38.

Gallina, A., Pollock, C.L., Vieira, T.M., Ivanova, T.D., Garland, S.J., 2016a. Between-day reliability of triceps surae responses to standing perturbations in people post-stroke and healthy controls: A high-density surface EMG investigation. Gait Posture 44, 103–109.

Gallina, A., Vieira, T., 2015. Territory and fiber orientation of vastus medialis motor units: A Surface electromyography investigation. Muscle Nerve 52, 1057–1065.

Gazzoni, M., Celadon, N., Mastrapasqua, D., Paleari, M., Margaria, V., Ariano, P., 2014. Quantifying Forearm Muscle Activity during Wrist and Finger Movements by Means of Multi-Channel Electromyography. PLoS One 9, e109943.

Gootzen, T.H.J.M., Vingerhoets, D.J.M., Stegeman, D.F., 1992. A study of motor unit structure by means of scanning EMG. Muscle Nerve 15, 349-357.

Herrmann, U., Flanders, M., 1998. Directional tuning of single motor units. J. Neurosci. 18, 8402–16.

Hodson-Tole, E.F., Wakeling, J.M., 2007. Variations in motor unit recruitment patterns occur within and between muscles in the running rat (Rattus norvegicus). J. Exp. Biol. 210, 2333–45.

Holt, G., Nunn, T., Allen, R.A., Forrester, A.W., Gregori, A., 2008. Variation of the Vastus Medialis Obliquus Insertion and its Relevance to Minimally Invasive Total Knee Arthroplasty. J. Arthroplasty 23, 600–604.

Huang, C., Chen, X., Cao, S., Zhang, X., 2016. Muscle-tendon units localization and activation level analysis based on high-density surface EMG array and NMF algorithm. J. Neural Eng. 13, 1–9.

Hug, F., Turpin, N. a, Dorel, S., Guével, A., 2012. Smoothing of electromyographic signals can influence the number of extracted muscle synergies. Clin. Neurophysiol. 123, 1895–6.

Jolliffe IT (1986). *Principal Component Analysis and Factor Analysis*, in Principal Component Analysis. Springer Series in Statistics. Springer, New York, NY.

Laine, C.M., Martinez-valdes, E., Falla, D., Mayer, F., Farina, D., 2015. Motor Neuron Pools of Synergistic Thigh Muscles Share Most of Their Synaptic Input. J Neurosci 35, 12207–12216.

Lambert-Shirzad, N., Van der Loos, H.F.M., 2017. On identifying kinematic and muscle synergies: a comparison of matrix factorization methods using experimental data from the healthy population. J. Neurophysiol. 117, 290–302.

Lee, D.D., Seung, H.S., 1999. Learning the parts of objects by non-negative matrix factorization. Nature 401, 788–91.

Lin, F., Wang, G., Koh, J.L., Hendrix, R.W., Zhang, L., 2004. In vivo and Noninvasive Three-Dimensional Patellar Tracking Induced by Individual Heads of Quadriceps. Med. Sci. Sport. Exerc. 36, 93–101.

Masani, K., Popovic, M., Nakazawa, K., Kouzaki, M., Nozaki, D., 2003. Importance of Body Sway Velocity Information in Controlling Ankle Extensor Activities During Quiet Stance. J. Neurophysiol. 90, 3774–3782.

Muceli, S., Jiang, N., Farina, D., 2013. Extracting Signals Robust to Electrode Number and Shift for Online Simultaneous and Proportional Myoelectric Control by Factorization Algorithms. IEEE Trans. Neural Syst. Rehabil. Eng 22, 623-633.

Negro, F., Holobar, A., Farina, D., 2009. Fluctuations in isometric muscle force can be described by one linear projection of low-frequency components of motor unit discharge rates. J. Physiol. 24, 5925–5938.

Peeler, J., Cooper, J., Porter, M.M., Thliveris, J.A., Anderson, J.E., 2005. Structural parameters of the vastus medialis muscle. Clin. Anat. 18, 281–289.

Rainoldi, A., Falla, D., Mellor, R., Bennell, K., Hodges, P., 2008. Myoelectric manifestations of fatigue in vastus lateralis, medialis obliquus and medialis longus muscles. J. Electromyogr. Kinesiol. 18, 1032–1037.

Rodriguez-Falces, J., Negro, F., Gonzalez-Izal, M., Farina, D., 2013. Spatial distribution of surface action potentials generated by individual motor units in the human biceps brachii muscle. J. Electromyogr. Kinesiol. 23, 766–77.

Roeleveld, K., Stegeman, D.F., Vingerhoets, H.M., Van Oosterom, A., 1997. Motor unit potential contribution to surface electromyography. Acta Physiol. Scand. 160, 175–183.

Smith, T.O., Nichols, R., Harle, D., 2009. Do the Vastus Medialis Obliquus and Vastus Medialis Longus Really Exist? A Systematic Review. Clin. Anat. 199, 183–199.

Staudenmann, D., Kingma, I., Daffertshofer, A., Stegeman, D.F., van Dieen, J.H., 2009. Heterogeneity of muscle activation in relation to force direction: A multi-channel surface electromyography study on the triceps surae muscle. J. Electromyogr. Kinesiol. 19, 882–895.

Staudenmann, D., Stegeman, D.F., van Dieën, J.H., 2013a. Redundancy or heterogeneity in the electric activity of the biceps brachii muscle? Added value of PCA-processed multi-channel EMG muscle

activation estimates in a parallel-fibered muscle. J. Electromyogr. Kinesiol. 23, 892–8.

Tenan, M.S., Hackney, A.C., Griffin, L., 2016. Entrainment of vastus medialis complex activity differs between genders. Muscle and Nerve 53, 633–640.

Tenan, M.S., Peng, Y.-L., Hackney, A.C., Griffin, L., 2013. Menstrual cycle mediates vastus medialis and vastus medialis oblique muscle activity. Med. Sci. Sports Exerc. 45, 2151–7.

Travnik, L., Djordjevič, S., Rozman, S., Hribernik, M., Dahmane, R., 2013. Muscles within muscles: a tensiomyographic and histochemical analysis of the normal human vastus medialis longus and vastus medialis obliquus muscles. J. Anat. 222, 580–7.

Tresch, M.C., Cheung, V.C.K., d'Avella, A., 2006. Matrix factorization algorithms for the identification of muscle synergies: evaluation on simulated and experimental data sets. J. Neurophysiol. 95, 2199–212.

Vieira, T., Botter, A., Minetto, M.A., Hodson-Tole, E.F., 2015. Spatial variation of compound muscle action potentials across human gastrocnemius medialis. J Neurophysiol 114, 1617–1627.

Vieira, T.M.M., Merletti, R., Mesin, L., 2010a. Automatic segmentation of surface EMG images: Improving the estimation of neuromuscular activity. J. Biomech. 43, 2149–2158.

Vieira, T.M.M., Windhorst, U., Merletti, R., 2010b. Is the stabilization of quiet upright stance in humans driven by synchronized modulations of the activity of medial and lateral gastrocnemius muscles ? J. Appl. Physiol. 85–97.

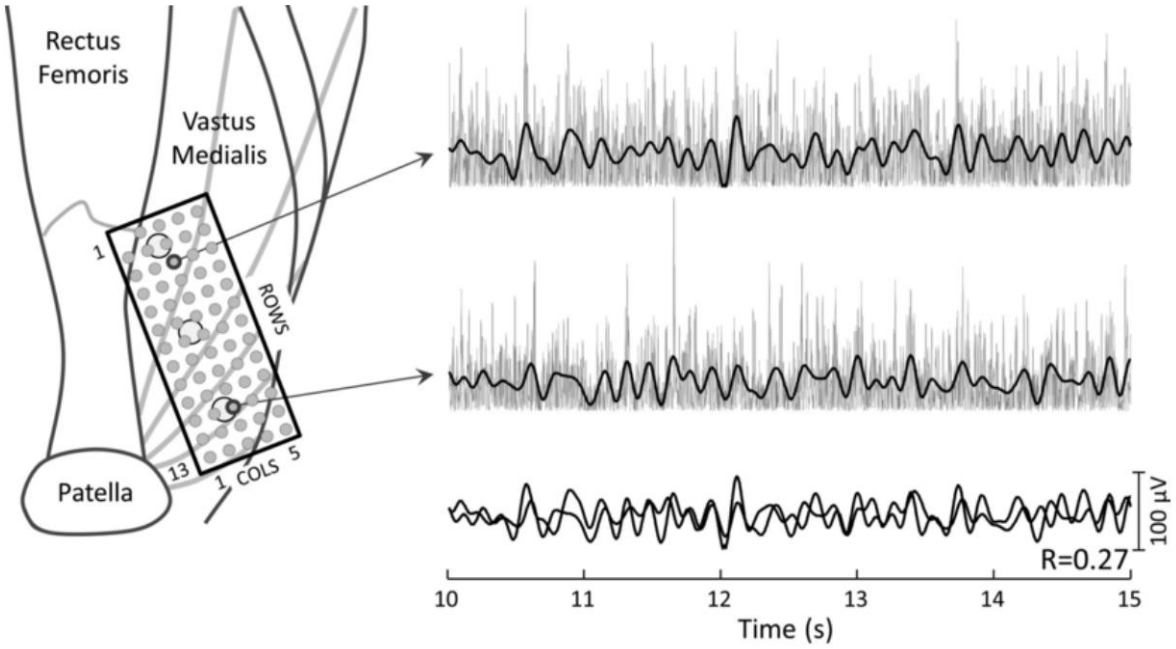
Wilson, N.A., Sheehan, F.T., 2010. Dynamic in vivo quadriceps lines-of-action. J. Biomech. 43, 2106–2113.

TABLE:

Table 1: Variance explained with PCA and NMF for EMG envelopes obtained with different cut-off frequencies. Values are averaged across participants.

	PCA1	PCA2	PCA3	PCA4	PCA5	NMF1	NMF2	NMF3	NMF4	NMF5
2Hz	77.0	89.6	93.9	95.9	97.0	77.0	88.6	90.1	90.8	91.3
4Hz	75.0	88.2	92.9	95.2	96.5	75.0	88.1	91.2	91.9	92.6
6Hz	72.9	87.6	92.6	94.9	96.4	72.9	87.6	92.0	93.1	93.6
8Hz	70.9	87.3	92.7	95.0	96.5	70.9	87.3	92.4	94.1	94.6
12Hz	69.5	87.0	92.7	95.2	96.6	69.5	87.0	92.6	94.7	95.6
20Hz	68.0	86.2	92.2	94.8	96.3	68.0	86.2	92.1	94.5	95.6
50Hz	63.8	82.6	89.1	92.5	94.6	63.8	82.6	89.0	92.2	94.1

509 **FIGURES:**



510

511 Figure 1: Left: Position of the electrode grid; the three large circles depict innervation zones identified in
512 different regions of the VM. Right (top two traces): Example of rectified EMG signals from two
513 electrodes 64 mm apart and the corresponding EMG envelope (thick black lines). Bottom trace: The two
514 envelopes are overlapped. The Pearson correlation coefficient between the two envelopes is 0.23,
515 meaning that the amount of common signal between the two locations is minimal.

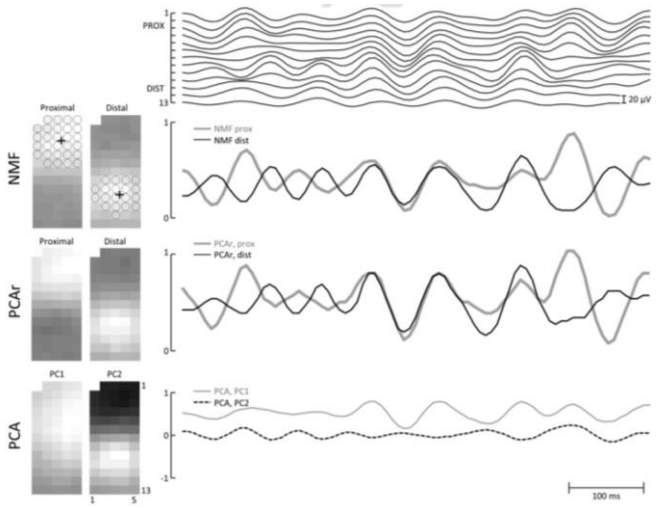


Figure 2: Top right: EMG envelopes from one column of the grid; their resulting factorization is shown below. Left: Spatial weights calculated with NMF, PCA and PCAr; the black circles identify the channels considered for the proximal and distal cluster, and the cross is their position. Right bottom: Temporal scores calculated with NMF, PCA and PCAr. Note the similarity between NMF and PCAr for both spatial weights and temporal scores.

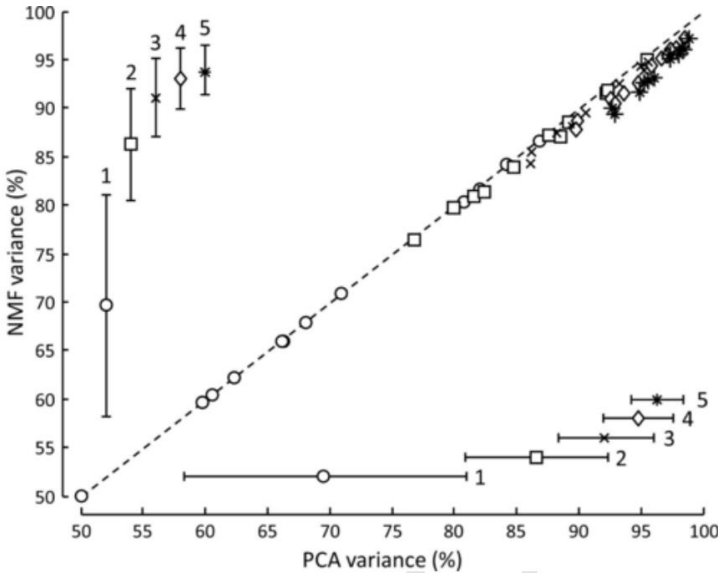


Figure 3: Correlation between the percentage of variance explained by PCA and NMF using one (circle) to five (asterisk) factors for all 12 participants. The dotted line is the bisecting line. Data presented on the left and on the bottom represent the average and standard deviation of NMF and PCA variance across all participants, respectively, for 1-5 factors.

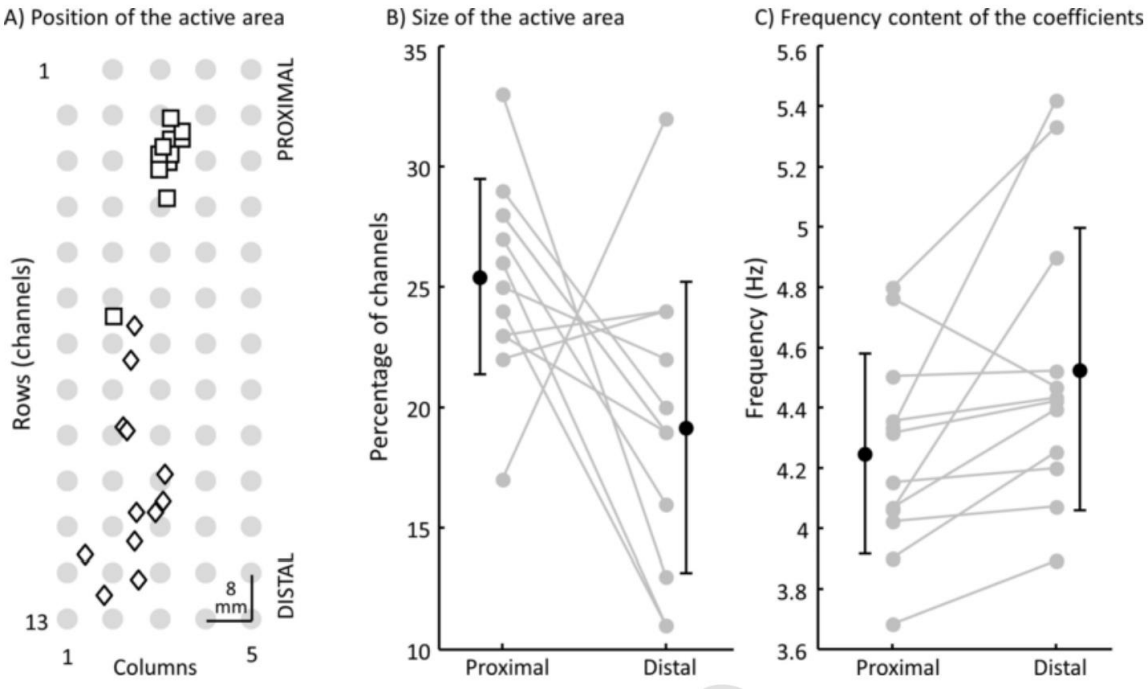


Figure 4: A) Location of the active area of the NMF factors for each of the 12 participants (squares, proximal; diamonds, distal). B) Size of the active area of each NMF factor (percentage of channels in the proximal and distal cluster). C) Median frequency of the fluctuation of the temporal score of the NMF coefficients.