Journal of Electromyography and Kinesiology 21 (2011) 811–818

Contents lists available at ScienceDirect

Journal of Electromyography and Kinesiology

journal homepage: www.elsevier.com/locate/jelekin

EMG-based muscle fatigue assessment during dynamic contractions using principal component analysis

Daniel R. Rogers [a,*], Dawn T. MacIsaac [b]

[a] Institute of Biomedical Engineering, Department of Electrical and Computer Engineering, University of New Brunswick, Fredericton, New Brunswick, Canada
[b] Department of Electrical and Computer Engineering, Faculty of Computer Science, University of New Brunswick, Canada

ARTICLE INFO

Article history:
Received 1 September 2010
Received in revised form 10 May 2011
Accepted 10 May 2011

Keywords:
EMG
Muscle fatigue
Fatigue index
Dynamic contraction
Principal component analysis

ABSTRACT

A novel approach to fatigue assessment during dynamic contractions was proposed which projected multiple surface myoelectric parameters onto the vector connecting the temporal start and end points in feature-space in order to extract the long-term trend information. The proposed end to end (ETE) projection was compared to traditional principal component analysis (PCA) as well as neural-network implementations of linear (LPCA) and non-linear PCA (NLPCA). Nine healthy participants completed two repetitions of fatigue tests during isometric, cyclic and random fatiguing contractions of the biceps brachii. The fatigue assessments were evaluated in terms of a modified sensitivity to variability ratio (SVR) and each method used a set of time-domain and frequency-domain features which maximized the SVR. It was shown that there was no statistical difference among ETE, PCA and LPCA ($p > 0.99$) and that all three outperformed NLPCA ($p < 0.0022$). Future work will include a broader comparison of these methods to other new and established fatigue indices.

1. Introduction

The assessment of muscle fatigue based on surface electromyograhy (sEMG) has been a prominent area of biomedical research since (Piper, 1912) first reported a "slowing" of the signal during a sustained contraction. This spreading of the sEMG in time results in a compression of the signal's power spectrum towards lower frequencies. Spectral parameters such as the mean power frequency (MF) or median frequency (MDF) have since been studied extensively as a means of tracking the progression of fatigue over time (Stulen and De Luca, 1981; De Luca, 1984). During sub-maximal contractions, amplitude measures of the sEMG such as average rectified value (ARV) and root mean square (RMS) value have generally been shown to increase with fatigue due to decreased muscle fiber propagation velocity (MFPV), increased motor unit (MU) firing rate, increased MU recruitment and increased MU synchronization (Lowery and O'Malley, 2003). (Dimitrova and Dimitrov, 2003) noted however, that the literature does contain some contradictory results regarding amplitude parameters during fatigue that can be explained by the effects of fatigue-induced changes in the shape of the intracellular action potential (decreased amplitude, increased negative after-potentials). These effects were shown to be dependant on the electrode-fiber distance which is inherently different for different muscles and subjects.

Although both amplitude and spectral parameters have been shown to reflect changes in the sEMG during isometric, constant force, fatiguing contractions, during dynamic contractions, where force and joint-angle vary with time, the non-stationary nature of the sEMG causes increased variability in these parameters. This is due to the recruitment and de-recruitment of active motor units in the vicinity of the detecting electrodes, the time-varying spatial filter as the muscle changes shape as well as the movement of the innervation zone relative to the surface electrodes, collectively termed dynamic factors. Thus, a robust fatigue index must detect the long-term changes in the sEMG while remaining insensitive to the short-term changes caused by dynamic factors.

A multivariable approach to fatigue assessment has recently been recommended (MacIsaac et al., 2006; Mesin et al, 2008; Rogers and MacIsaac, 2009) with the suggestion that multiple features provide a more complete set of information than any single parameter. (MacIsaac et al., 2006) showed that a multivariable mapping index (MI) tracked fatigue with less variability than MF. MI and a more generalized version of the index, (GMI) (Rogers and MacIsaac, 2009) both use baseline data to train a feed-forward neural network having one hidden layer to output an estimate of fatigue. This network architecture closely resembles the mapping portion of a non-linear multivariate data analysis technique proposed by Kramer (1991), termed non-linear principal component analysis (NLPCA). Instead of using baseline data to train a neural network, Kramer uses an auto-associative architecture which connects the output of a mapping network directly to the input of a

* Corresponding author.
E-mail address: dan.rogers@unb.ca (D.R. Rogers).

1050-6411/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jelekin.2011.05.002

demapping network in order to reproduce the inputs at the output. Following training, the mapping–demapping network can be split so the mapping network can be used on its own. Kramer showed that if linear transfer functions are used in all neurons, the resulting linear principal component analysis (LPCA) is equivalent to PCA computed by traditional Eigen decomposition.

This work applied traditional PCA as well as Kramer's LPCA and NLPCA along with a fourth novel multivariate analysis technique termed end-to-end projection (ETE) in order to compare each method's ability to extract long-term changes in the sEMG in the context of tracking muscle fatigue. All of these approaches make no assumption about the linearity of the features with time and none require separate training data thus overcoming the two drawbacks of MI cited by MacIsaac. Due to space limitations, this study focused only on multi-variable feature projection methods leaving to a future study the comparison of the best of these methods to existing fatigue indices.

2. Methodology

Fatigue indices are often evaluated using a sensitivity-based metric such as percent change from initial to final to value. While sensitivity to fatigue is clearly important, such a metric gives no indication of the variability of the index over time. A signal-to-noise ratio (SNR) such as the one proposed by (MacIsaac et al., 2006) considers sensitivity in its numerator, and variability in its denominator. A modified version of the metric was proposed by (Rogers and MacIsaac, 2009) as given by (1) where $\hat{I}$ is the two-segment piece-wise linear line of best fit of the index I and P is the number of estimates (i.e. temporal points). Although a strictly linear trend with time is sometimes observed, the choice of a piece-wise linear fit more closely matches empirical observations as compared to other non-linear fits such as quadratic or exponential.

This metric compares the long-term trend (i.e. the net change of the index) to the variability about that trend and thus can be considered a sensitivity-to-variability (SVR) metric. High sensitivity in I (estimated by the range of $\hat{I}$) increases the SVR and variability in I (i.e. difference between I and $\hat{I}$) lowers the SVR.

$$\mathrm{SVR} = \frac{R}{E_{\mathrm{RMS}}} = \frac{\max_p(\hat{I}_p) - \min_p(\hat{I}_p)}{\sqrt{\frac{1}{P}\sum_{p=1}^{P}(I_p - \hat{I}_p)^2}} \quad (1)$$

Consider a set of N myoelectric features, where each feature is represented by a time-series of estimates, one every t_s seconds for the duration of a contraction lasting T seconds. At a given point in time, the vector containing the value of each feature represents a single point in N-dimensional feature-space. Over the entire contraction the series of points creates a scatter in feature-space.

For stationary sEMG, the resulting features show no statistical change with time. In feature-space the scatter of points lies within a cluster whose location and spread are given by the statistics of the features. Specifically, the centroid of the cluster is at the point $U = \{\mu_1, \mu_2, \ldots, \mu_N\}^T$ where μ_i is the true mean of the i^{th} feature and T denotes matrix transpose (not to be confused with the signal duration T). The spread of the cluster in the i^{th} dimension is given by the variance of the i^{th} feature (which is not necessarily independent of other features). The cluster represents the region in feature-space where the features are statistically most likely to lie.

For non-stationary sEMG (such as from a fatiguing contraction), the underlying means of the features change with time, denoted U_t instead of U. This corresponds to a migration of the cluster's centroid through feature-space beginning at the point U_0 (the feature mean at $t = 0$) and ending at the point U_T (the feature mean at $t = T$). As the true means of the features change with time, the scatter of points is stretched out along the trajectory of the centroid. Whereas changes in the feature means affect the location of the centroid, variations in variance or higher order statistics affect how spread out the scatter of points is distributed along the centroid trajectory. This trajectory is the global long-term trend in the features and the scatter about the trajectory is the noise about that trend. Thus, an alternative interpretation of the SVR can be given as the ratio between the 'length' of trajectory to the 'width' of variation of the feature-space scatter of points.

The trajectory through feature-space is not necessarily linear. The linear case results only if the underlying trend with time is linear for all features or if the trends are non-linear but equal or opposite, regardless of the variability about the trend. In such a case, traditional PCA could be used to accurately capture the linear feature-space trajectory. PCA projects data onto eigenvectors of the data's covariance or correlation matrix. The eigenvectors are an orthogonal set of vectors which capture the dimensions of maximum variance in feature-space. In the case where the length of the centroid trajectory is greater than the maximum variance about the trajectory, the trajectory is the dimension of maximum variance and the eigenvector corresponding to the largest eigenvalue (and thus the first principal component) captures the temporal trend in the features.

It is generally not the case that sEMG features exhibit a linear trend in feature-space. (Merletti et al., 1991) described sEMG amplitude features as often exhibiting non-linear "dome-shaped" trends with fatigue whereas spectral parameters maintained linear or curvilinear decreasing trends. (Dimitrova and Dimitrov, 2003) noted that amplitude features are affected by changes in the intracellular action potential which cause the time-course of the features to vary depending on the electrode-fiber distance. Furthermore, dynamic factors introduce non-stationarities which may be reflected differently in different features. Thus, in general a non-linear migration of the feature centroid through feature-space should be assumed. In this case, Kramer's NLPCA may be better suited to capture the feature-space trend.

This work proposes a novel approach to capture the long-term trend in multiple myoelectric features by projecting the data onto the vector which connects the points U_0 and U_T in feature-space. This approach preserves the temporal information so that the fatigue index which results from this end to end (ETE) projection always represents the global change of the features with respect to their initial and final values.

As a practical example, Fig. 1 shows ARV and MF features recorded during a fatiguing dynamic contraction (normalized to have unit maximum). Fig. 2 shows the feature-space trajectory of these two features as well as the estimated ETE vector. Projection of the data onto the ETE vector results in the fatigue index shown in Fig. 2. Comparing $\mathrm{SVR}_{\mathrm{ETE}} = 15.9$ to the individual features, $\mathrm{SVR}_{\mathrm{ARV}} = 11.8$ and $\mathrm{SVR}_{\mathrm{MF}} = 6.5$, a substantial increase in SVR (i.e. decrease in variability) is achieved by considering the long-term trend information in two features as compared to either individually.

2.1. Fatigue test protocol

This work used the same sEMG data that (Rogers and MacIsaac, 2009) used to demonstrate a generalized form of MacIsaac's MI fatigue index by showing that a general function could be trained using the training data from other individuals. The sEMG was collected using the testing protocol and disc-pulley apparatus described by (MacIsaac et al, 2006) which consisted of a large disc which was free to rotate about its centre and was mounted within a rigid frame. The participant was seated with the elbow near the axis of rotation of the disc with the forearm extended toward the edge. A weight supported by a cable was mounted along the

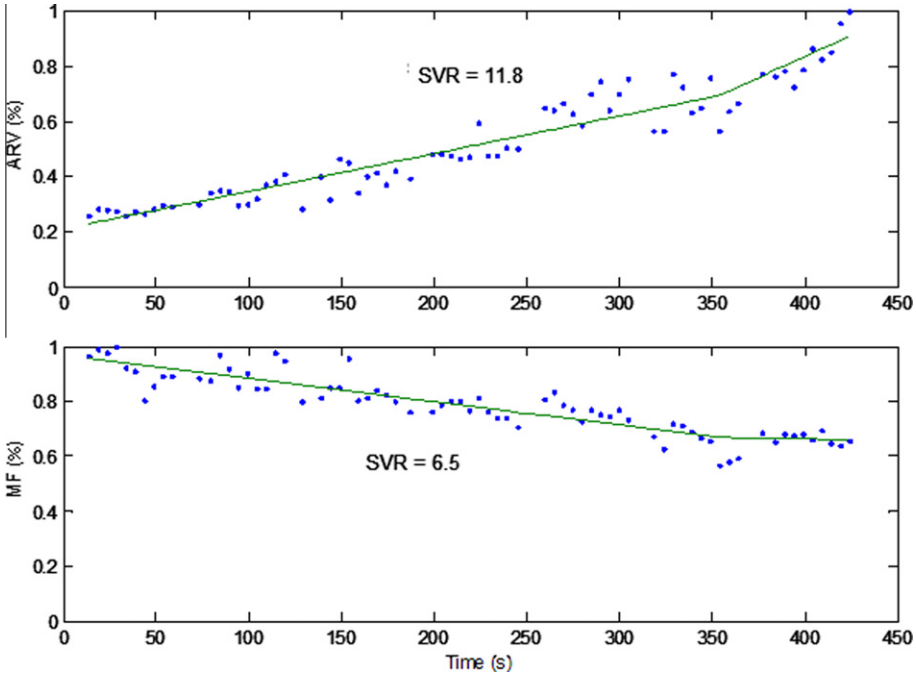

Fig. 1. Average Rectified Value (ARV) and Mean Frequency (MF) calculated from sEMG of a dynamic fatiguing contraction. Features are normalized to have a maximum of one.

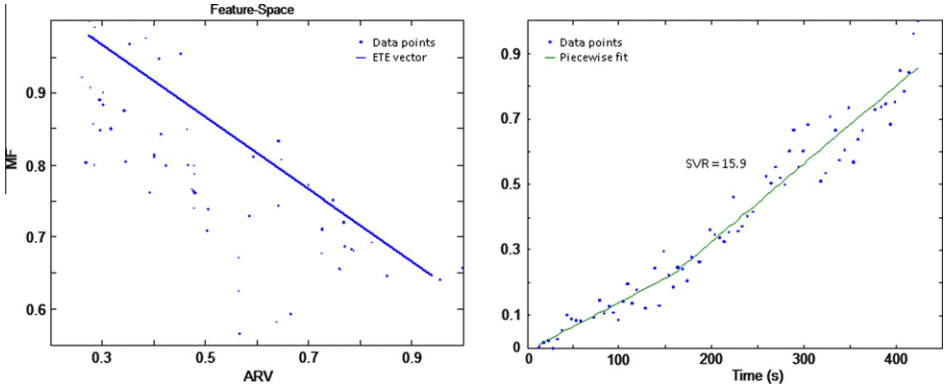

Fig. 2. Left: Feature-space of sEMG features shown in Error! Reference source not found. and the estimated end to end (ETE) vector which shows the feature-space trajectory with time. Right: Fatigue index resulting from projection of data onto ETE vector shown on left. Also shown is the piece-wise line of best fit used for calculating SVR.

circumference of the disc providing resistance to elbow flexion that was perpendicular to the forearm regardless of joint angle.

With approval from the ethics board at the University of New Brunswick, two sets of data were recorded under static (i.e. isometric), cyclic and random contraction conditions (six sessions in total) from nine healthy human participants (four males and five females aged 28 ± 8 years old) who each gave signed, informed consent. At least 3 days rest was allowed between each of the six sessions in order to ensure that the muscle was fully recovered from the previous session. Surface EMG was collected from the right biceps brachii as the participant performed the prescribed contraction until exhaustion. Each participant used a weight of approximately 30% of maximum voluntary contraction at 130° joint angle (where 180° is full extension), determined prior to the first session. An ergometer attached at the disc's axis of rotation was used to record the joint angle as well as to provide visual feedback to the participant in the form of an onscreen tracking target. By following as closely as possible a second, computer-generated target, the user performed the desired contraction. Static contractions were held at 90° joint angle and dynamic contractions consisted of elbow flexion and extension between 50° and 130° moving at approximately 32°/s. For random contractions, the computer-generated target changed direction between 0.75 and 2 s according to a uniform distribution.

In order to confirm a fatiguing process, dynamic (cyclic and random) contractions were interrupted for the first ten seconds of each minute. During this time an isometric contraction was held at 90° joint angle and a MF calculation was obtained. A statistically significant negative slope of these readings was used to confirm that the muscle had fatigued.

2.2. Data acquisition

Surface EMG was acquired using an adhesive eightchannel Ag–AgCl electrode strip with 5 mm spacing which was placed along the midline of the biceps brachii, proximal to the muscle belly and approximately parallel to the muscle fibers. Signals collected by the electrode were processed by a Prima EMG16 multi-channel acquisition system described by (Martin and MacIsaac, 2006) with differential gains between 1000 and 10,000. Seven channels of

single-differential sEMG were then sent to the computer for digital storage using a National Instruments DAQ-6024E PCMCIA data acquisition card sampling at 1024 Hz.

Although only a single channel of sEMG was required for fatigue estimation, the ideal placement of a single electrode pair is difficult to predict for a given individual (MacIsaac et al, 2001; Dimitrova et al, 2009). The linear array was used as a means to ensure at least one channel was located away from both the innervation zone and musculotendon region.

2.3. Data processing

Data processing was done offline using Matlab®. The intermittent periods of isometric contraction were examined to confirm a fatiguing process. For static contractions, the first ten seconds of each minute was used and for dynamic contractions, the joint angle data collected from the ergometer was used to identify the intermittent isometric segments. Fatigue was confirmed using a regression analysis of MF values, estimated per segment, each calculated from a periodogram calculated by averaging 0.5 s, 50% overlapped Hamming-windowed epochs.

The sEMG was then segmented as follows: static and random data were divided into equal length, non-overlapping 5 s segments and cyclic data was segmented with one cycle per segment, beginning and ending at the minimum joint angle. Although the cyclic segments were different lengths in terms of the number of samples, the period of the cyclic contraction was approximately 5 s.

2.3.1. Myoelectric feature selection

Each sEMG segment was characterized by a set of eight features: four time-domain and four frequency-domain as given by (2). The time-domain features given by (3) are those described by (Hudgins et al, 1993) and include ARV, the number of zeros crossings (ZC), the number of slope sign changes (SSC) and the waveform length (WL) as given by (4–7), respectively where $x_p[k]$ is the kth sample of the p^{th} sEMG segment having K samples.

$$F_p = \left\{ \begin{matrix} F_p^{\mathrm{TDF}} \\ F_p^{\mathrm{FDF}} \end{matrix} \right\} \tag{2}$$

where F_p^{TDF} and F_p^{FDF} are the time-domain and frequency-domain features respectively, given by Eqs. (3) and (8), respectively.

$$F_p^{\mathrm{TDF}} = \{\mathrm{ARV}_p, \mathrm{ZC}_p, \mathrm{SSC}_p, \mathrm{WL}_p\}^T \tag{3}$$

where

$$\mathrm{ARV}_p = \frac{1}{K}\sum_{k=1}^{K} |x_p[k]| \tag{4}$$

$$\mathrm{ZC}_p = \frac{N_{\mathrm{ZC}}}{K} \tag{5}$$

where *NZC* is incremented if

$$sign(x_p[k] \cdot x_p[k+1]) < 0 \tag{5a}$$

while

$$|x_p[k] - x_p[k+1]| > \varepsilon_1 \tag{5b}$$

where ε_1 is threshold used to reduce over-counting due to noise.

$$\mathrm{SSC}_p = \frac{N_{\mathrm{SSC}}}{K} \tag{6}$$

where N_{SSC} is incremented if

$$\begin{matrix} (x_p[\mathrm{k}-1] < x_p[\mathrm{k}] \mathrm{and} x_p[\mathrm{k}] > x_p[\mathrm{k}+1]) \\ \mathrm{or} \\ (x_p[\mathrm{k}-1] > x_p[\mathrm{k}] \mathrm{and} x_p[\mathrm{k}] < x_p[\mathrm{k}+1]) \end{matrix} \tag{6a}$$

while

$$|x_p[\mathrm{k}-1] - x_p[\mathrm{k}]| > \varepsilon_2 or |x_|pk] - x_p[\mathrm{k}+1]]| > \varepsilon_2 \tag{6b}$$

where ε_2 is threshold used to reduce over-counting due to noise.

$$\mathrm{WL}_p = \frac{1}{K}\sum_{k=2}^{K} |x_p[\mathrm{k}] - x_p[[\mathrm{k}-1]| \tag{7}$$

Frequency-domain features given by (8) include spectral power (SP), MF, MDF and 95% bandwidth (BW) given by (9, 12), respectively, where $\Phi_p(f)$ is the power spectral density of the pth sEMG segment calculated by averaging 0.5 s, 50% overlapped, Hamming-windowed epochs, and M is number of frequency bins in $\Phi_p(f)$. For sampling frequency 1024 Hz and 0.5 s epochs, $M = 256$.

$$F_p^{FDF} = \{\mathrm{SP}_p, \mathrm{MF}_p, \mathrm{MDF}_p, \mathrm{BW}_p\}^T \tag{8}$$

$$\mathrm{SP}_p = \sum_{m=1}^{M} \Phi_p(f_m) \tag{9}$$

$$\mathrm{MF}_p = \frac{\sum_{m=1}^{M} f_m \Phi_p(f_m)}{\mathrm{SP}_p} \tag{10}$$

$$\mathrm{MDF}_p = f_K \tag{11}$$

such that

$$\sum_{m=1}^{K} \Phi_p(f_m) = \sum_{m=K+1}^{M} \Phi_p(f_m) \tag{11a}$$

$$\mathrm{BW}_p = f_H - f_L \tag{12}$$

such that

$$\sum_{m=L}^{H} \Phi_p(f_m) = 0.95 \cdot SP_p \tag{12a}$$

and

$$\sum_{m=1}^{L} \Phi_p(f_m) = \sum_{m=H}^{M} \Phi_p(f_m) \tag{12b}$$

Each feature of a given contraction was normalized such that the maximum value was unity. Although all the chosen features are known to vary with fatigue, they almost certainly contain redundant information. It is possible that a particular feature might not contribute any new information about the long-term trend but would still contribute noise in feature-space. This work sought to find the optimum subset of the eight features which best captured the long-term trend (i.e. had the highest SVR). Thus, for each method (PCA, LPCA, NLPCA, ETE), a brute force approach was taken by evaluating a fatigue estimate using all 255 possible feature subsets. Using all of the collected data, the feature subset which achieved the highest SVR when averaged across participants, contraction types and trials was chosen as the optimal feature set to be used for comparison to the other methods.

2.3.2. PCA

PCA is a mathematical operation which defines a new set of orthogonal axes in feature space in order to maximize the variance of the data along each axis. PCA projects the data onto the N

eigenvectors of the features' covariance matrix or correlation matrix. According to (Joliffe, 2002), the covariance matrix is appropriate for normalized data (i.e. data whose features have similar variance) which is the case in this work. If the direction of maximum variance in feature-space is due to the temporal trajectory (i.e. in the case of a linear trajectory), the first principal component will track fatigue.

2.3.3. LPCA and NLPCA

Feed-forward auto-associative neural networks were used according to (Kramer, 1991). These are two stage networks such that the outputs of a mapping stage are fed directly to the inputs of a demapping stage. The mapping portion consists of an input layer, hidden layer and bottleneck layer such that the number of neurons in the bottleneck layer is less than the number of inputs, creating a lower dimensional representation of the N input features. In this work the bottleneck layer had a single neuron and the hidden layer had N neurons. N varied from one to eight during feature set optimization depending on the number of features in the feature subset being tested. The demapping network reverses the transformation of the mapping network by reconstructing the N input features from the lower dimensional (in this case one-dimensional) representation of the bottleneck layer. The demapping network consisted of a hidden layer and an output layer both having N neurons. The transfer function of neurons in the hidden layers were logistic sigmoids for NLPCA and linear for LPCA. Transfer functions for the bottleneck and output layers were linear in all cases.

A backpropagation technique was used to train the mapping–demapping networks using the input features as the desired outputs. Following training, the network was split apart and only the mapping portion was used to map the N features to a single output from the bottleneck layer.

2.3.4. End to end projection

The ETE vector D which connects points U_0 and U_T in feature-space is simply the difference given by (13). Since these points represent the statistical mean of the features at $t = 0$ and $t = \mathrm{T}$, respectively, they are unknown in practice and were estimated by averaging the first three and last three data points, respectively, as given by (14). Vector D can be thought of as an axis in feature-space which spans the distance between the start and end points of the contraction. The ETE fatigue index, I_{ETE}, is simply the projection of the set of feature vectors, F, onto D as given by (15). Eq. (15) is the scalar product of each data point and the vector D. The value of the scalar is the length of the projection of the feature vector F along the vector D. Thus the fatigue index, I_{ETE}, tracks the temporal evolution of the features along the path from the start to the end of the contraction.

$$\mathrm{D} = U_T - U_0 \tag{13}$$

$$\hat{\mathrm{D}} = \frac{1}{3}\sum_{p=P-2}^{P} F_p - \frac{1}{3}\sum_{p=1}^{3} F_p \tag{14}$$

$$I_{\mathrm{ETE}} = \hat{\mathrm{D}}^T F \tag{15}$$

2.3.5. Assessing fatigue

Since seven channels of differential sEMG were collected, each channel was processed as described above yielding a fatigue index for each channel. The channel with the highest SVR was used as the fatigue estimate for the given fatigue index. Although this means that for a given contraction record, the channel used for ETE may be different from the channel used for PCA, LPCA or NLPCA, this allows each index to perform as well as possible without favouring any particular index. For visual comparison among indices of fatigue, all indices were scaled such that the estimated trend, $\hat{I}$ spanned the range [0, 1].

2.4. Statistical analysis

To determine if there were any statistical differences among the fatigue indices, a three-way ANOVA with repeated measures was used considering the fatigue index, contraction type and trial as factors. The ANOVA was followed by Bonferroni-corrected pairwise comparisons. Statistical analysis was done offline using Minitab® version 15.1.

3. Results

All nine participants completed the six fatiguing contractions. Endurance times were as follows: $T = 409 \pm 151$ s, $T_{\min} = 255$ s, $T_{\max} = 922$ s. For all contraction records, fatigue was confirmed by a regression analysis which indicated a statistically significant ($p < 0.05$) negative slope of the MF values calculated from the intermittent isometric segments.

3.1. Feature selection

The brute force approach examining all 255 feature subsets was done by averaging the SVR of fatigue estimates over all participants, contraction conditions and trials. Table 1 shows the mean SVR values for different feature subsets of interest. The optimum feature set was defined as the subset which had the highest average SVR. For PCA and LPCA the optimum feature set was given by (16) and for NLPCA and ETE by (17) and (18), respectively.

$$F_{\mathrm{PCA}}^{\mathrm{OPT}} = \{\mathrm{ARV}, \mathrm{ZC}, \mathrm{SSC}, \mathrm{MF}, \mathrm{MDF}, \mathrm{BW}\}^T \tag{16}$$

$$F_{\mathrm{NLPCA}}^{\mathrm{OPT}} = \{\mathrm{ARV}, \mathrm{ZC}, \mathrm{MF}, \mathrm{BW}\}^T \tag{17}$$

$$F_{\mathrm{ETE}}^{\mathrm{OPT}} = \{\mathrm{ARV}, \mathrm{ZC}, \mathrm{SSC}, \mathrm{MF}, \mathrm{BW}\}^T \tag{18}$$

3.2. Comparing fatigue indices

Fig. 3 shows the normalized myoelectric features from trial two for participant six during a random contraction and Fig. 4 shows the four normalized fatigue indices for the same participant and trial under static, cyclic and random contraction conditions. These results were typical of all participants. Fig. 5 shows the SVR values for all nine participants and both trials under static, cyclic and random contraction conditions. The three-way ANOVA showed no significant difference between trials ($p > 0.91$) and no interaction between trials and contraction conditions ($p > 0.95$), trials and fatigue indices ($p > 0.73$) or trials, conditions and indices ($p > 0.72$). Trial was therefore removed from the model and a two-way ANOVA was done considering fatigue indices and contraction conditions. In this case contraction condition showed no effect ($p > 0.57$) and there was no interaction with fatigue indices ($p > 0.16$). Since contraction condition had no effect, it was removed from the analysis

Table 1
SVR of fatigue estimates averaged over all participants, trials and contraction conditions for different methods and feature sets (N represents the # input features).

Method	Optimum	TDF + FDF	TDF	FDF
PCA	12.6 ± 5.4 ($N = 6$)	9.2 ± 3.9	8.6 ± 3.9	9.7 ± 4.0
LPCA	12.6 ± 5.4 ($N = 6$)	9.2 ± 3.9	8.6 ± 3.9	9.7 ± 4.0
NLPCA	11.3 ± 4.7 ($N = 4$)	8.8 ± 3.9	8.7 ± 4.0	9.3 ± 4.0
ETE	13.0 ± 5.2 ($N = 5$)	10.3 ± 3.9	9.6 ± 3.6	10.5 ± 4.1

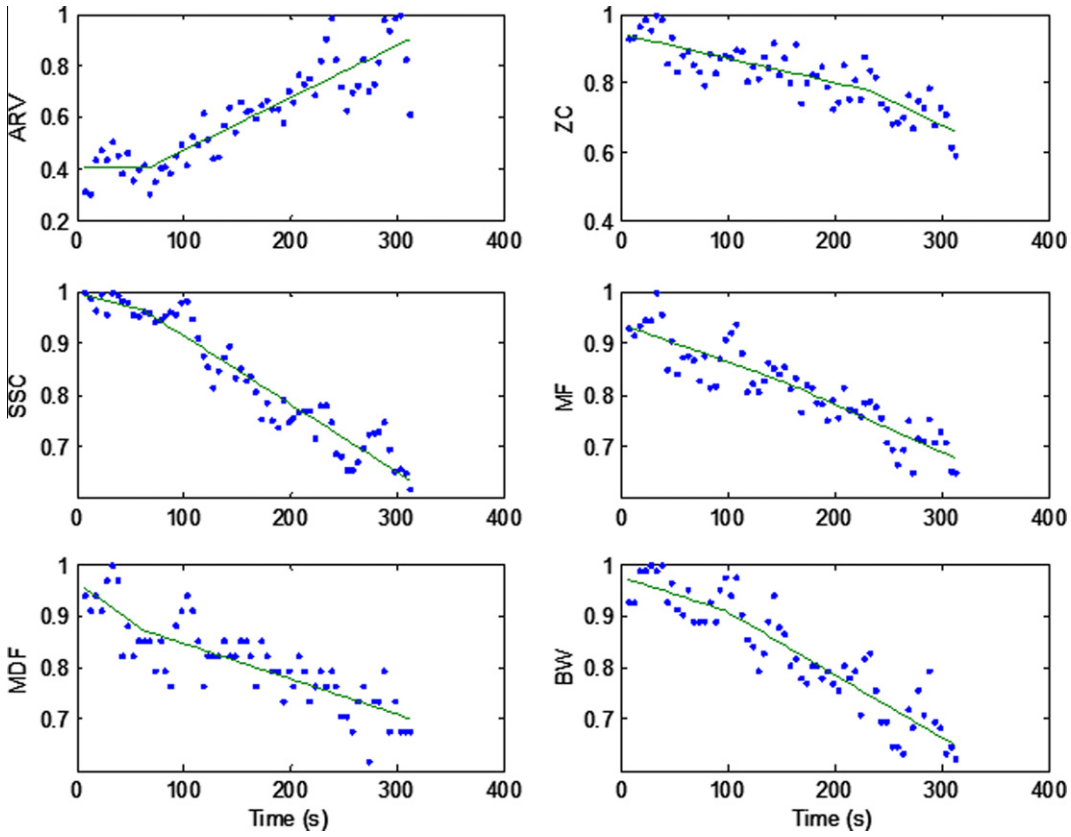

Fig. 3. Normalized myoelectric features for random contraction of participant six, trial two, channel seven. This channel was optimal for all indices.

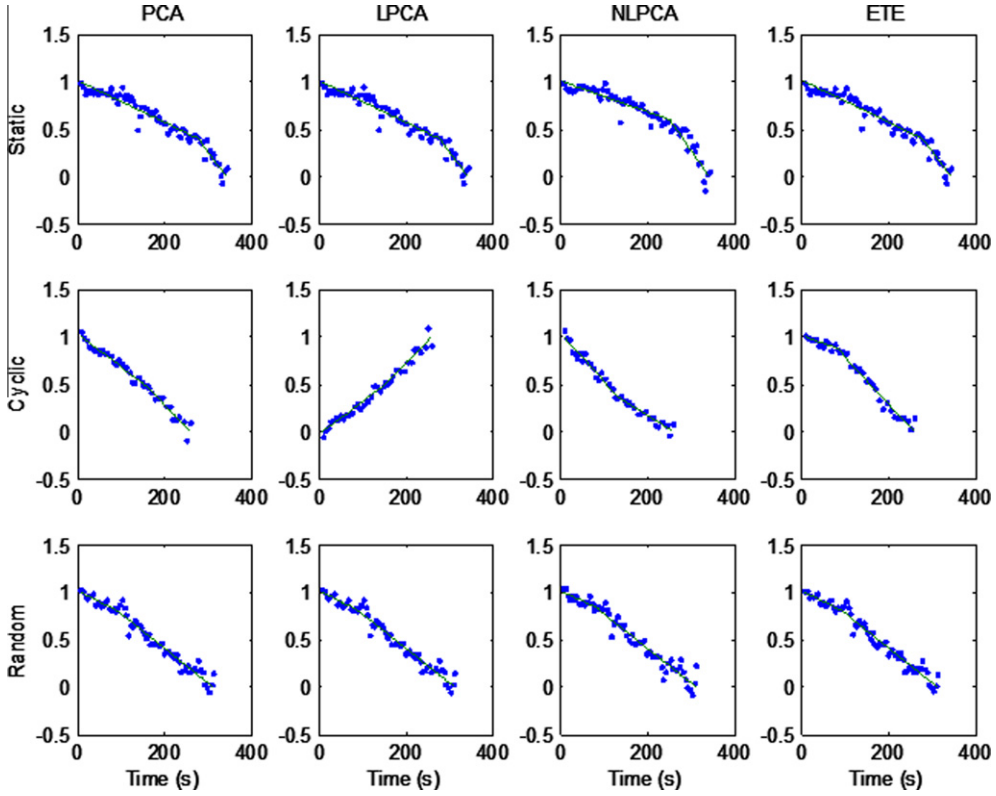

Fig. 4. Fatigue estimates for participant six, trial two, under static, cyclic and random contraction conditions. Vertical axes were normalized for visual comparison.

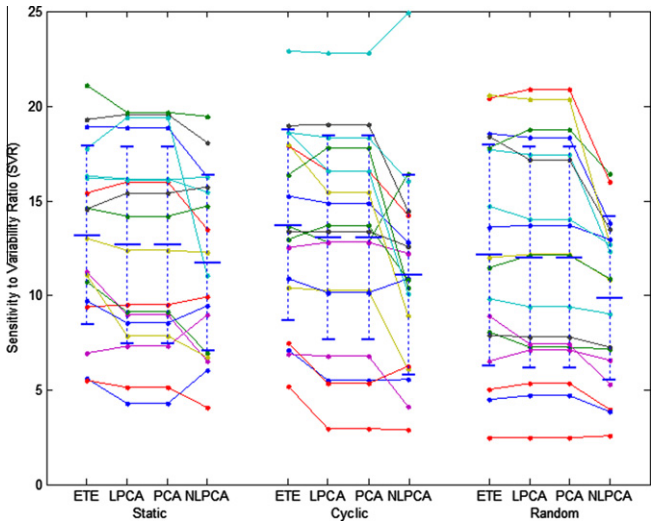

Fig. 5. SVR values for both trials of all nine participants under static, cyclic and random contraction conditions (bars indicate mean and one standard deviation). Indices are shown left to right in descending order of mean SVR averaged across participants, trials and conditions.

and a third ANOVA was done considering only fatigue index. This analysis showed a statistical difference among indices ($p < 0.0001$). The p-values associated with the Bonferroni-corrected pairwise comparisons showed that ETE, PCA and LPCA all significantly outperformed NLPCA ($p < 0.0022$) but there was no difference among ETE, PCA and LPCA.

4. Discussion

All of the optimum feature sets given by (16–18) included ARV, ZC, MF and BW indicating that these features contain the most trend information. Also, all optimum sets excluded WL and SP from the full feature set given by (2). Both WL and SP are indirect ways of measuring the sEMG amplitude; it is likely that ARV, which measures amplitude directly, captures as much or more of the amplitude information thereby making WL and SP redundant.

The use of optimal feature sets for each index was done in order to maximize the performance of each index. Although the lack of an independent data set on which to conduct the optimization may lead to overly optimistic results, this approach avoids favouring any particular index.

The time-domain and frequency-domain features were chosen because they capture sEMG amplitude and spectral information which is known to vary with fatigue. They are also well established in the literature and are relatively easy to compute. While an optimum subset of those features was determined for this study, this does not constitute a thorough feature-set optimization. The methods investigated constitute a general approach to capturing the feature-space trajectory for extracting long-term trend information and could be used with features other than those explored in this study (e.g. autoregressive coefficients, normalized spectral moments, etc).

In the study of muscle fatigue using sEMG, it is implicitly assumed that the observable changes in muscle performance and the changes in the sEMG are the result of the underlying physiological processes of muscle fatigue. Because the individual effects on the sEMG of the many interrelated physiological parameters remain unresolved, it is difficult to draw any physiological conclusions from the observed myoelectric features or indices. This work, therefore, focused only on the myoelectric manifestations of fatigue without attempting attribute these changes to any physiological mechanisms.

The linear electrode array was used to guarantee at least one channel that was sufficiently far from both the innervation zone and musculotendon region. As discussed in detail in (Rogers and MacIsaac, 2009), the selection of a channel for analysis proved non-trivial as results were inconsistent across channels and fatigue indices. The choice of the best sEMG channel means that for a particular participant and contraction condition, the channel used for ETE may be different from that used for PCA, LPCA, or NLPCA. Although this approach may yield overly optimistic SVR values, it gives each index the opportunity to do its best without favouring any particular index. Future work will examine the consistency and repeatability of myoelectric parameters from multi-channel sEMG.

Statistical analysis showed that the linear projection methods (PCA, LPCA and ETE) all significantly outperformed NLPCA. This is a strong indication that the feature-space trajectories were nearly linear. Although this conclusion is worthy of note, it cannot be generalized beyond the scope of the muscle and force level that was examined in this study. Even in this study, given the relatively small sample size (nine participants), power in the statistical tests is low. Caution should therefore be employed when making an assumption about the linearity of trajectory.

As to why NLPCA was not able to approximate the linear case as well as PCA and LPCA, it is likely due to the non-optimal projection demonstrated by (Malthouse, 1998). The NLPCA method is constrained to a continuous projection function which causes non-orthogonal projection of the data onto the feature-space curve (even if the curve is close to linear).

Inspection of the SVR values in Fig. 5 shows that PCA and LPCA gave identical results in all cases. This supports the assertion made by (Kramer, 1991) that the two methods are mathematically equivalent.

The ETE vector in feature-space represents the net change from the start to the end of the contraction. Since the end points U_0 and U_T represent the statistical means of the features, estimation of these points by sample-averaging is inherently inexact. This represents a trade-off between the accuracy of the fit in the case of a linear feature-space trajectory and the interpretability of the results in the non-linear case. In the case of a purely linear feature-space trajectory, the eigen decomposition used by PCA does a more accurate job of capturing the trajectory than ETE because it considers all of the data points. For a non-linear feature-space trajectory, the eigenvector corresponding to the largest eigenvalue does not necessarily represent the temporal evolution of the features and therefore the resulting projection may not intuitively be interpreted as capturing the long-term trend. ETE conversely, although slightly less accurate in the purely linear case, retains the interpretation of representing the long-term trend in the data (possibly with some scale compression) regardless of the linearity of the feature-space trajectory.

Needing to have all of the data in order to generate a fatigue index would seem to preclude the real-time application of any of these techniques in monitoring fatigue. However, if the fatigue index is re-calculated as new data arrive, the resulting index will relate the myoelectric features' current state to their initial state. This is not different from established fatigue indices such as MF or Dimitrov's normalized spectral moments. For example, an initial MF calculation of 70 Hz gives no indication of whether or not the muscle is fatigued; it is only after a relative change in MF such as a decrease to 60 Hz that indicates a manifestation of fatigue. Thus, it is not the absolute value of a parameter that is of interest but the relative change over time. Although the absolute values of previously calculated fatigue estimates would change as vector projections are re-calculated, the relative change in the parameters

would still be captured as it occurs. Whereas PCA and especially ETE can be recalculated quickly, depending on the available computing capacity the re-training of neural networks may not meet real-time requirements.

5. Conclusion

Traditional PCA along with neural-network implementations of linear and non-linear PCA were compared to a novel method for the extraction of a global long-term trend in multiple myoelectric features in the context of muscle fatigue assessment. The novel method projected sEMG features onto the estimated ETE vector in feature-space in order to extract the global long-term trend information. The ETE projection represents the relative global change of the parameters from initial to final value.

It was shown that similar mixtures of time-domain and frequency-domain features yielded the best performance when the variability of fatigue estimates was assessed in terms of a piecewise linear SVR metric. It was shown that the linear approaches (PCA, LPCA and ETE) outperformed the non-linear NLPCA.

Future work will include a broader comparison of these approaches to other new and established fatigue indices.

Acknowledgment

The authors would like to thank Katelyn Kent for overseeing the experimental work which was carried out at the Institute of Biomedical Engineering, University of New Brunswick. This work was supported in part by grants from the Natural Sciences and Engineering Research Council of Canada.

References

De Luca CJ. Myoelectrical manifestations of localized muscular fatigue in humans. CRC Crit Rev Biomed Eng 1984;11:251–79.

Dimitrova NA, Dimitrov GV. Interpretation of EMG changes with fatigue: facts, pitfalls and fallacies. J Electromyogr Kinesiol 2003;13:13–36.

Dimitrova NA, Arabadzhiev TI, Hogrel J-Y, Dimitrov GV. Fatigue analysis of interference EMG signals obtained from *biceps brachii* during isometric voluntary contractions at various force levels. J Electromyogr Kinesiol 2009;19:252–8.

Hudgins B, Parker PA, Scott RN. A new strategy for multi-function myoelectric control. IEEE Trans Biomed Eng 1993;40(1):82–94.

Joliffe IT. Principal Component Analysis. second ed. New York: Springer; 2002. pp. 21–26.

Lowery MM, O'Malley MJ. Analysis and simulation of changes in EMG amplitude during high-level fatiguing contractions. IEEE Trans Biomed Eng 2003;50(9):1052–62.

Kramer MA. Nonlinear principal component analysis using auto-associative neural networks. AIChE J 1991;37:233–43.

MacIsaac DT, Parker PA, Englehart KB, Rogers DR. Fatigue estimation with a multivariable myoelectric mapping function. IEEE Trans Biomed Eng 2006;53(4):694–700.

MacIsaac D, Parker PA, Scott RN. The short-time Fourier transform and muscle fatigue assessment in dynamic contractions. J Electromyograp Kinesiol 2001;11:439–49.

Malthouse EC. Limitations of nonlinear PCA as performed with generic neural networks. IEEE Trans Neural Network 1998;9(1):165–73.

Martin S, MacIsaac D. Innervation zone shift with changes in joint angle in the brachial biceps. J Electromyograp Kinesiol 2006;16:144–8.

Merletti R, Lo Conte LR, Orizio C. Indices of muscle fatigue. J Electromyograp Kinesiol 1991;1:20–3.

Mesin L, Cescon C, Gazzoni M, Merletti R. A new method to estimate myoelectric manifestations of muscle fatigue. Proc XVII Int. Soc. Electromyogr and Kinesiol Congr 2008; Niagara Falls, Canada.

Piper H. Electrophysillogie Muschlicke Muskeln. Basel, Switzerland: Verlag von Julius; 1912.

Rogers DR, MacIsaac DT, Training a multivariable myoelectric mapping function to estimate fatigue. J Electromyograp Kinesiol (2009), doi:10.1016/j.jelekin.2009.11.001.

Stulen F, De Luca CJ. Frequency parameters of the myoelectric signal as a measure of muscle conduction velocity, IEEE Transactions on Biomedical Engineering 1981; BME-28: 512–523.

Daniel R. Rogers received the B.Sc.E. degree from the University of New Brunswick, Fredericton, NB, Canada in 2005. He is currently a research associate and doctoral student with the Institute of Biomedical Engineering, Department of Electrical and Computer Engineering, University of New Brunswick. His research interests include muscle fatigue assessment during dynamic contractions and control systems for powered lower limb prostheses.

Dawn T. MacIsaac received the Ph.D. degree from the University of New Brunswick, Fredericton, NB, Canada in 2004. She received the M.Sc.E. degree in electrical and computer engineering from the University of New Brunswick in 1999, the B.Eng. degree in electrical engineering from McMaster University, Hamilton ON, Canada in 1996, the B.Ed. degree from Queen's University, Kingston, ON, Canada in 1991 and the B.PE. degree from the Department of Physical Education, McMaster University in 1990. She is currently working at the University of New Brunswick as an Associate Professor in the Department of Electrical and Computer Engineering and the Faculty of Computer Science. Her research interests include dynamic myoelectric parameter estimation, muscle fatigue, artificial neural networks, time-frequency analysis and software engineering. She is also Coordinator of Teaching and Learning Services in the Center for Enhanced Teaching and Learning.