Accuracy and learning curves of inexperienced observers for manual segmentation of electromyograms

Acurácia e curva de aprendizado de observadores inexperientes para segmentação manual de eletromiogramas

Arthur de Sá Ferreira[a], Fernando Silva Guimarães[b], Manuel Armando Ribeiro Magalhães[c], Regina Coeli Souza e Silva[d]

[a] Physical therapist, doctorate in Biomedical Engineering, professor of Laboratory of Computational Simulation and Modeling in Rehabilitation, Post-Graduation Program of Rehabilitation Science, Augusto Motta University Center (Unisuam), Rio de Janeiro, RJ - Brazil, e-mail: arthur_sf@ig.com.br

[b] Physical therapist, doctorate in Clinical Medicine, professor of Laboratory of Cardiovascular and Respiratory Performance, Post-Graduation Program of Rehabilitation Science, Augusto Motta University Center (Unisuam), Rio de Janeiro, RJ - Brazil, e-mail: fguimaraes_pg@hotmail.com

[c] Physical therapist, graduation student, Laboratory of Computational Simulation and Modeling in Rehabilitation, Post-Graduation Program of Rehabilitation Science, Augusto Motta University Center (Unisuam), Rio de Janeiro, RJ - Brazil, e-mail: armando.fisioterapia@gmail.com

[d] Physical therapist, graduation student, Laboratory of Computational Simulation and Modeling in Rehabilitation, Post-Graduation Program of Rehabilitation Science, Augusto Motta University Center (Unisuam), Rio de Janeiro, RJ - Brazil, e-mail: coeli.unisuam@yahoo.com.br

Abstract

Introduction: The shape-varying format of surface electromyograms introduces errors in the detection of contraction events. **Objective**: To investigate the accuracy and learning curves of inexperienced observers to detect the quantity of contraction events in surface electromyograms. **Materials and methods**: Six observers performed manual segmentation in 1200 shape-varying waveforms simulated using a phenomenological model with variable events, smooth changes in amplitude, marked on-off timing, and variable signal-to-noise ratio (0-39 dB). Segmentation was organized in four sessions with 15 blocks of 20 signals each. Accuracy and learning curves were modeled per block by linear and power regression models and tested for difference among sessions. Cut-off values of signal-to-noise ratio for optimal manual segmentation were also estimated. **Results**: The accuracy curve showed no significant linear trend throughout blocks and no difference among

sessions 1-2-3-4 (87% [85; 89], 87% [85; 89], 87% [85; 89], 87% [81; 88]; p = 0.691). Accuracy was low for detection of 1 event (AUC = 0.40; sensitivity = 44%; specificity = 43%; cut-off = 12.9 dB) but was high and affected by the signal-to-noise ratio for detection of two events (AUC = 0.82; sensitivity = 77%; specificity = 76%; cut-off = 7.0 dB). The learning curve showed a significant power regression (p < 0.001) with decreasing values of learning percentages (time duration to complete the task) among sessions 1-2-3-4 (86.5% [68; 94], 76% [68; 91], 62% [38; 77], and 57% [52; 75]; p = 0.002). **Conclusion**: Inexperienced observers exhibit high, not trainable accuracy and a practice-dependent shortening in the time spent to detect the quantity of contraction events in simulated surface electromyograms.

Keyword: Electromyography. Computer simulation. Muscle.

Resumo

Introdução*: A forma de onda variável do eletromiograma de superfície introduz erros na detecção de eventos de contração. **Objetivo**: Investigar as curvas de acurácia e aprendizagem de observadores inexperientes para detectar a quantidade de eventos em eletromiogramas de superfície. **Materiais e métodos**: Seis observadores realizaram segmentação manual em 1.200 sinais com formas de onda variável simulada usando modelo fenomenológico, com mudanças suaves em amplitude, tempos início-fim marcados, e relação sinal-ruído variável (0-39 dB). A segmentação foi organizada em quatro sessões, com 15 blocos de 20 sinais. As curvas de acurácia e aprendizagem foram modeladas por bloco utilizando modelos de regressão linear e de potência, e foram testados quanto à diferença entre as sessões. Valores de corte de relação sinal-ruído ideais para segmentação manual também foram estimadas. **Resultados**: A curva de acurácia não mostrou tendência linear significativa e não houve diferença entre as sessões 1-2-3-4 (87% [85; 89], 87% [85; 89], 87% [85; 89], 87% [81; 88], p = 0,691). A acurácia foi baixa para a detecção de um evento (AUC = 0,40; sensibilidade = 44%; especificidade = 43%; limiar = 12,9 dB), mas foi elevada e afetada pela relação sinal-ruído para a detecção de dois eventos (AUC = 0,82; sensibilidade = 77%, especificidade = 76%; limiar = 7,0 dB). A curva de aprendizagem mostrou uma regressão de potência significativa (p < 0,001) com valores decrescentes de percentagens de aprendizagem (tempo de duração para completar a tarefa) entre sessões 1-2-3-4 (86,5% [68; 94], 76% [68; 91], 62% [38; 77] e 57% [52; 75], p = 0,002). **Conclusão**: Observadores inexperientes apresentam alta acurácia não treinável e uma redução prática-dependente do tempo gasto para detectar a quantidade de eventos de contração em eletromiogramas de superfície simulados.*

Palavras-chave*: Eletromiografia. Simulação computacional. Muscular.*

Introduction

The surface electromyogram (SEMG) exhibits a shape-varying waveform related to neural strategies for motor units recruitment during muscle contractions (1, 2). A comprehensive understanding of muscle's physiology depends on an accurate estimation of parameters related to its activity, which in turn relies on how accurately the SEMG is segmented into contraction events (3). Automated methods are fast and accurate for estimating on-off timing of events exhibiting approximately constant high amplitude as obtained during maximal isometric voluntary contractions (4 - 7). However, they exhibit poor performance in case of shape-varying SEMG due to superposed

activation patterns of different movements (8), e.g. daily-living activities and dynamic sports activity. In such cases, manual segmentation by visual inspection can be used for screening SEMG to judge if the signal does or does not represent meaningful physiological activity (9). Despite its off-line, time-consuming characteristics (8, 9), manual segmentation is still performed because it provides highly accurate event detection (6). Accurate detection of events has important applications in movement sciences (3) and other fields such as estimation of preterm labor detection (10, 11) and tremor characterization (12). However, it is unknown how accurately observers execute this signal processing in shape-varying SEMG

and if training by repetition is sufficient to improve their accuracy.

To estimate the accuracy of observers for manual segmentation of SEMG, it is necessary that the analyzed signals present known features, i.e. quantity of contraction events, on-off timing, and signal-to-noise (SNR) ratio. Stochastic simulation of SEMG is of particular advantage in accuracy studies since the desired features can be controlled and arranged in a large number of combinations to yield gold-standard references. Several physical or phenomenological models have been developed to represent SEMG during sustained static contractions (13, 14), single bursts (5, 6, 15, 16), physiological phenomena (16, 17), and repetitive-movement tasks (18, 19) with application to studies for SEMG segmentation. The above-cited methods were used to assess errors for on-off timing of bursts but none of those studies investigated the accuracy for detecting the quantity of contraction events by manual segmentation.

The general hypothesis for knowledge acquisition by task repetition is that observers will perform progressively faster as they repeat the same procedure. The learning curve is a useful method to quantify the reduction in time to execute a procedure as a function of repetition (20, 21). Determination of learning and accuracy curves of observers performing manual segmentation of SEGM may contribute to establishing training parameters as well their accuracy level. Therefore, the aims of this study are: 1) to quantify the accuracy curve of inexperienced observers to detect the quantity of contraction events, and 2) to analyze their learning curves for manual segmentation of SEMG.

Methods

Subjects

Six inexperienced observers (4 women; 30 ± 14 years) were recruited from the graduation and post-graduation academic community. They studied surface electromyography during graduation and post-graduation courses but did not perform manual segmentation of signals previously to this study. All observers were informed about the procedures and gave their written consent. This study was approved by the institutional ethics committee before its execution (CAAE-0011.0.307.000-11).

Phenomenological method for simulation of shape-varying SEGM

The scheme for SEMG simulation was based on previous studies (5, 18, 19) and is depicted in Graph 1. The simulated SEMG(i) should present variable quantities of events, smooth changes in amplitude with marked on-off timing, and variable SNR. Therefore, the discrete raw SEMG(i) was represented by equations 1-2:

(1) $SEMG(i) = y(i) + e(i)$

(2) $y(i) = (i) \cdot \sum_{n=1}^{3} g_n(i) \cdot \prod_{n=1}^{3} s_n(i)$

where: $y(i)$ = noiseless SEMG for the ith sample (i = 1,2,…,2000); $e(i)$ = background noise modeled as a bandlimited (80-120 Hz, 1st order Butterworth filter) pseudorandom pattern; $r(i)$ = isometric contraction also modeled as a bandlimited Gaussian-distributed pseudorandom pattern with standard deviation σ_r; $g_n(I)$ = profiles of muscle activity modeled as n = 1, 2, 3 Gaussian functions with standard deviations σ_n and random amplitude factors in range [0.1; 1.0]; and $s_n(i)$ = on-off periods modeled as n square patterns with time support α_n and unitary amplitude. Power-line interference and motion artifact were not included since them could be satisfactorily removed before segmentation (22 - 25). The SNR ratio per event was calculated as the $10\log_{10}(\sigma_y^2/\sigma_e^2)$, where σ_y^2 and σ_e^2 represent calculated variances of $y(i)$ and $e(i)$, respectively. All signals were simulated with a sampling frequency of 1.0 kHz. No additional signal processing was performed in the SEMG(i) before manual segmentation. Notice that events independently simulated with random parameters may overlap in time, thus being considered as a single contraction event for comparison purposes (Graph 1).

Globally, 1.200 SEMG were synthesized using sets of uniformly distributed random values for σ_n (single events durations in range 50 to 150 ms), α_n (1 to 2.5), and SNR (0 to 39 dB). The duration of events was chosen to match those observed in tremor detection (12). The large range of SNR was chosen based on other studies (5) and to represent events that are either easy or difficult to find to distinguish performance of the observers. All simulated components – $r(i)$, $g_n(i)$, $s_n(i)$, and $e(i)$ – were stored as ASCII files in a database and are available upon request. Notice that the SNR does not depend on the duration of the simulated contraction event.
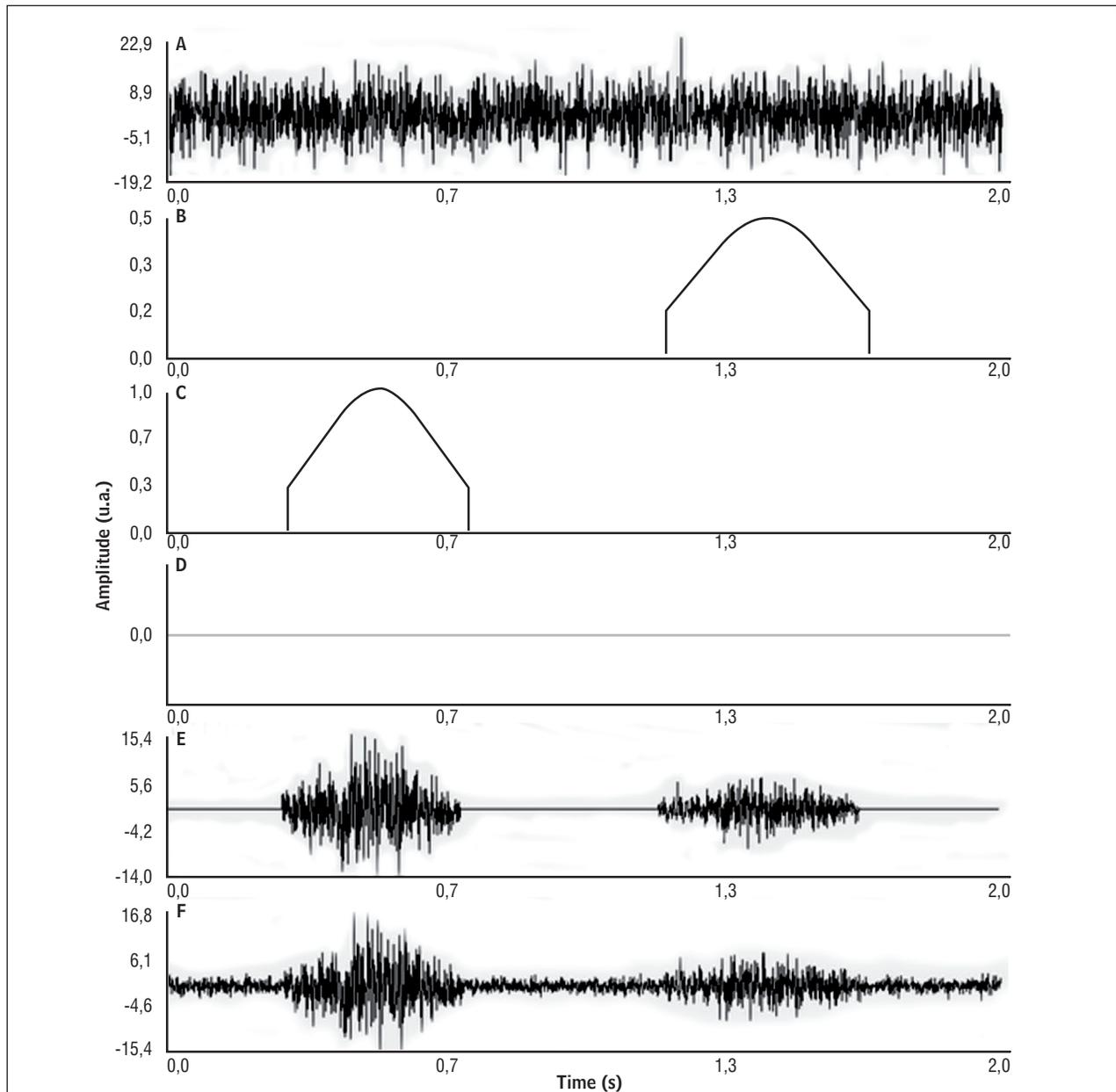
Graph 1 - Composition of the simulated surface electromyogram

Legend: From top to bottom: A = Gaussian pattern representing an isometric contraction; B-D = Gaussian patterns representing smooth muscle activity truncated by square patterns representing marked on-off muscle activity; E = Simulated surface electromyogram; F = Output surface electromyogram with additive noise.

Source: Research data.

Note: Notice that this signal (# 72 in database) was simulated with two contraction events and exhibited a signal-to-noise ratio equal to 13 dB and 7 dB for the first and second events, respectively.

Procedure for manual segmentation and computational resources

Manual segmentation was organized in four sessions, twice a week, with 15 blocks of 20 signals per session. Observers were advised to take the necessary time for analyses and were informed that signals might present up to three contraction events. In the beginning, observers selected the current block and ran an algorithm to open-read-display the signal in the sequence as stored in database. Only SEMG(i) was displayed in screen presenting with buttons and

movable cursors, making observers blinded to simulation patterns. They were instructed to accurately detect the quantity of events and to mark the corresponding button in the screen. In sequence, the algorithm displayed pairs of cursors for each detected event to allow observers to mark on-off timings for each event. The observer then closed the window and the next signal in the block was displayed. At the end of each block a pop-up window displayed the total time spent in manual segmentation for annotation in a paper worksheet (hh:mm:ss format). The quantity of contraction events and their respective on-off timing were digitally stored for comparative analysis per signal. A 5-minute rest was allowed between sequential blocks. In total, each session had approximately 2-3 hours including the rest period.

Two computers with the same configuration (Intel® Core 2 Duo, Windows® XP) were used in this study and observers used the same computer throughout the study. All algorithms for simulation and analysis were implemented in LabVIEW 8.0 (National Instruments, Texas, USA) and were fully automated.

Statistical analysis

For a given observer, signals were considered as correctly identified if the quantity of events marked by the observer matched the quantity estimated from the Boolean OR comparison of $s_n(i)$ signals (gold-standard of the activation pattern). Accuracy for quantification of events was computed as the proportion between the total of signals correctly identified to the total of signals per block. The accuracy curve per observer was modeled by a linear regression model (H_0: $\beta = 0$; intercept = 0) and tested for difference among sessions using the Friedman's test. Receiver-operating characteristic (ROC) curves (26) were used to determine the cut-off for SNR (continuous variable) to a successful detection (binary variable: correct = 1) per observer. The group-median area under the ROC curve (AUC), sensitivity, specificity, and cut-off values for SNR were estimated separately for signals simulated with 1 and 2 events (963 and 229 signals, respectively).

The time spent in manual segmentation per signal was estimated by dividing the time to analyze the block by the quantity of signals in the block (n = 20). The learning curve per observer was modeled by the Wright's

(1936) curve using the power regression analysis (H_0: $\beta = 0$; constant = 0). Learning percentages were tested for difference among sessions using the Friedman's test and Wilcoxon rank sum test for pairwise comparisons.

Statistical analyses were executed in LabVIEW 8.0 and SPSS 17 (SPSS Inc., Chicago, IL, USA). Descriptive values are shown as median [max; min]. Graphs display group-average values and error bars represent ± SEM. Statistical significance was considered at p < 0.05 (one-tailed) with the adjusted p-value based on the stepwise rejection Li's procedure (27). P-values were estimated by bootstrap procedure by using a Monte Carlo method with 800 samples.

Results

The accuracy curve showed no obvious trend throughout the blocks (Graph 2). No significant linear regressions on accuracy were obtained for session 1 (R = 0.289, R^2 = 0.084, p = 0.296), session 2 (R = 0.387, R^2 = 0.149, p = 0.155), session 3 (R = 0.033, R^2 = 0.001, p = 0.908), and session 4 (R = 0.332, R^2 = 0.111, p = 0.226) (Graph 3). No significant difference on accuracy was observed among sessions 1-2-3-4 (87% [85; 89], 87% [85; 89], 87% [85; 89], 87% [81; 88], respectively; p = 0.691).

ROC curves showed different behaviors considering the quantity of simulated events (Graph 3). For signals with 1 event, accuracy was low (AUC = 0.40) and was not affected by SNR (sensitivity = 44%; specificity = 43%; cut-off = 12.9 dB). For signals with two events, accuracy was high (AUC = 0.82) and was significantly affected by the SNR (sensitivity = 77%; specificity = 76%; cut-off = 7.0 dB).

The learning curve showed a "fast" decrease in time for segmentation of the first 10 blocks with small "peaks" occurring after blocks 15, 30, and 45 marking the beginning of each session (Graph 4). Significant power regressions were obtained for sessions 1 (R = 0.948, R^2 = 0.898), session 2 (R = 0.790, R^2 = 0.624), session 3 (R = 0.821, R^2 = 0.673), and session 4 (R = 0.758, R^2=0.574), all at p < 0.001. Decreasing values of learning percentages were observed among sessions 1-2-3-4 (86.5% [68; 94], 76% [68; 91], 62% [38; 77], and 57% [52; 75], respectively; p = 0.002) with significant differences among sessions 1-3 (p = 0.013), 1-4 (p = 0.013) and 2-3 (p = 0.016), but not significant between sessions 1-2 (p = 0.050), 2-4 (p = 0.033) and 3-4 (p = 0.420).

$y = -0.2589x + 89.127$ $\quad$ $y = -0.6875x + 101.87$ $\quad$ $y = -0.0536x + 86.147$ $\quad\quad$ $y = 0.4881x + 60.02$
$R^2 = 0.0835$ $\quad\quad\quad$ $R^2 = 0.14948$ $\quad\quad\quad$ $R^2 = 0.00106$ $\quad\quad\quad\quad$ $R^2 = 0.11055$
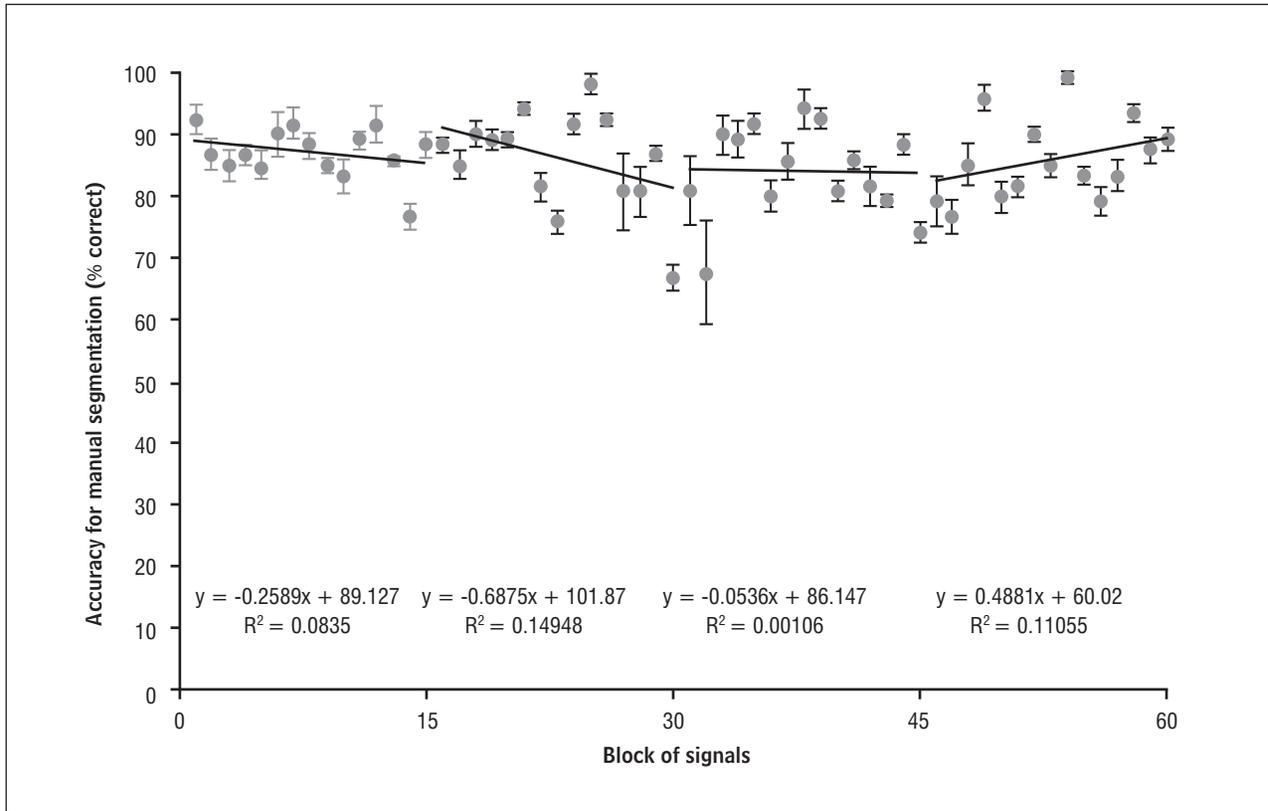
Graph 2 - Accuracy curve for manual segmentation of simulated surface electromyograms
Source: Research data.
Note: Values are presented as group-average per block. Error bars represent one standard error of mean.
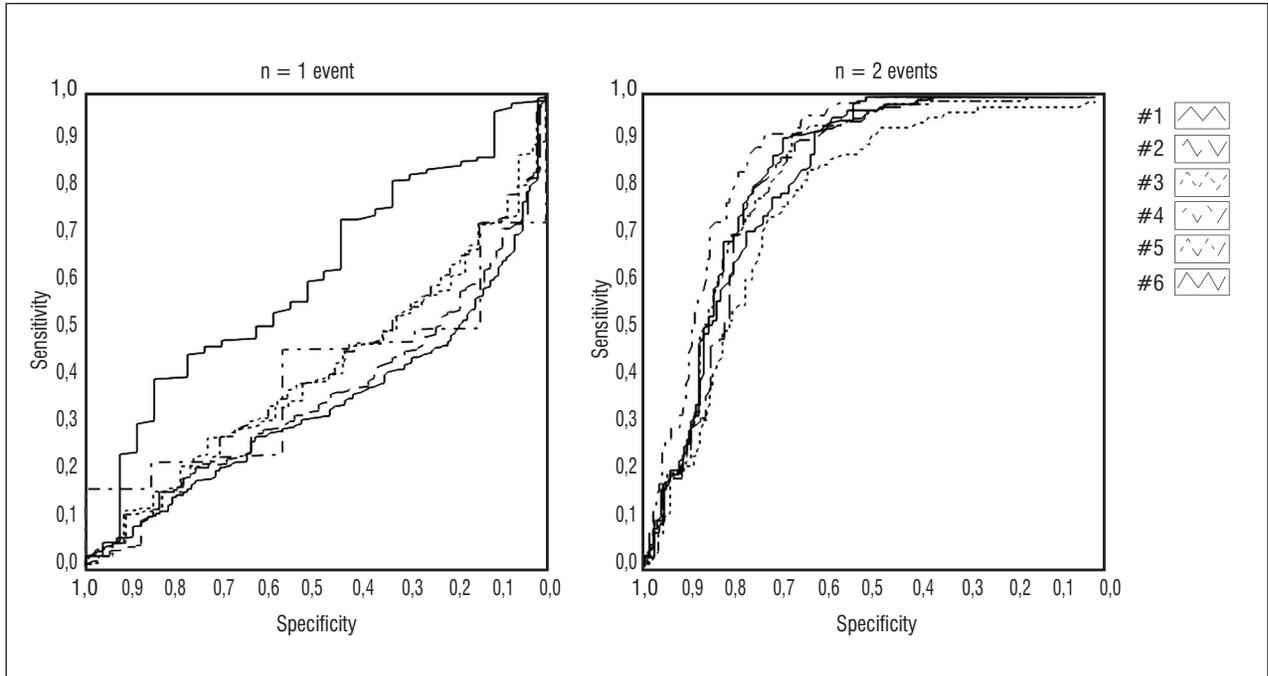
Graph 3 - Receiver operating characteristic curves for accuracy estimated for quantification of quantity of contraction events
Source: Research data.
Note: Lines represent each observer. Left: n =1 event; Right: n =2 events.
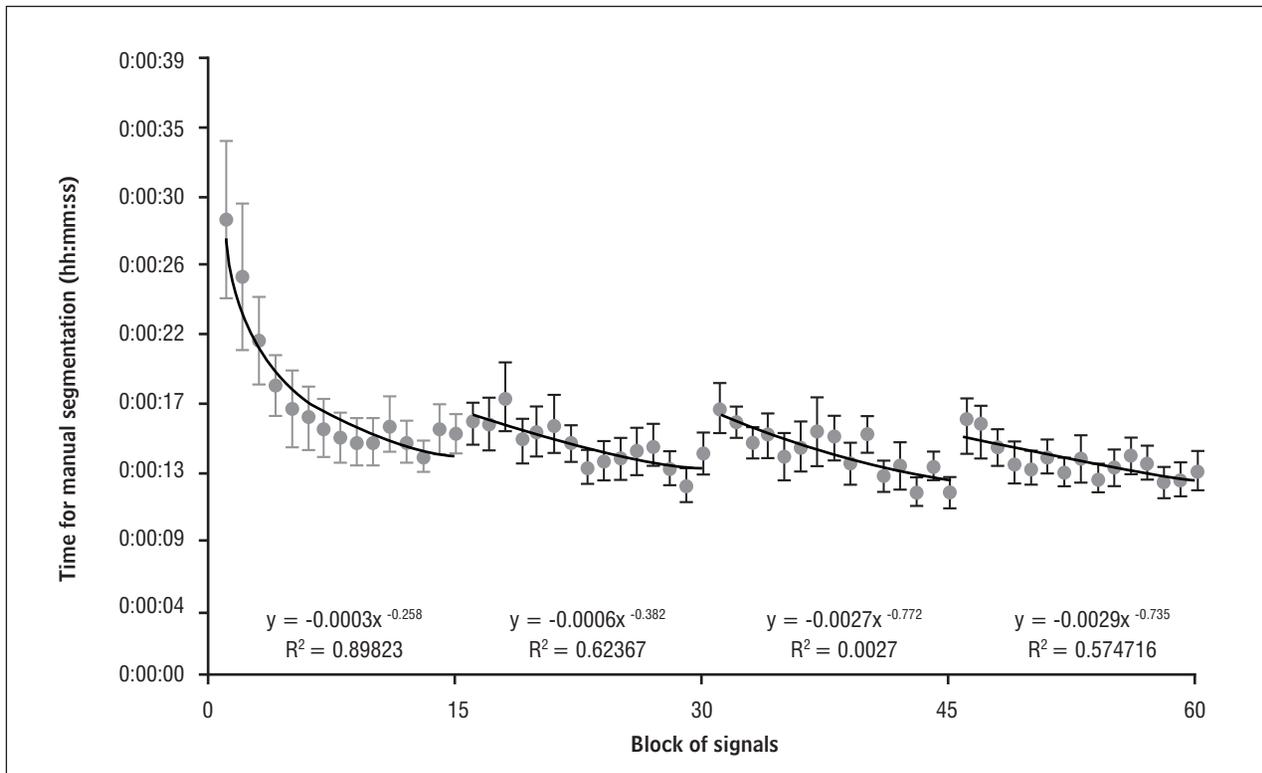
Graph 4 - Learning curve for manual segmentation of simulated surface electromyograms

Source: Research data.

Note: Values are presented as group-average per block (20 trials each block). Error bars represent one standard error of mean.

Discussion

This study quantified the accuracy of inexperienced observers to detect contraction events and described their learning curve for manual segmentation of SEMG. The main result is that high, stable accuracy for manual segmentation was obtained by simple repetition although accompanied by an increase in average speed for this signal processing. To the best of our knowledge, this is the first study to quantify the accuracy and learning curves of inexperienced observers for manual segmentation of SEMG. The level of accuracy obtained in this study between simulated signals and manual segmentation (85% to 89%) with variable SNR ratios is similar to that reported in a recent study, where a double-threshold method exhibited high agreement (87.5%) with manual segmentation of "clear periods" (unknown SNR) of muscle activity and inactivity in SEMG of patients with cervical spondylotic myelopathy during gait (28).

Detailed analysis of ROC curves showed that observers exhibited low statistical performance for detection of 1 event at a high SNR but performed better for detection of two events at a lower SNR. The results from the ROC curves with two events are considered the closest scenario to the real one, since the distinct behavior observed for signals simulated with one or two events can be explained by the experimental procedure. On one hand, observers were aware on the range of events and thus they may have been induced to select 'at least one' even if it was not clearly distinguished from baseline. On the other hand, for signals simulating two events the observers were more likely to miss events with smaller amplitudes. For comparison, the double-threshold method of Bonato et al. (5) exhibited high accuracy (> 95%) for the percentage of erroneous transitions in on-off timing for SNR above 10 dB. Also, several automated methods for on-off timing showed systematic degradation of accuracy with acceptable results at 6 dB but not at 3 dB or lower (12, 29). Therefore, the accuracy of manual segmentation obtained in this study is considered comparable to those exhibited by automated methods using simulated or real SEMG signals.

Observers were not aware about their accuracy from previous sessions. Such a lack of feedback concerning their performances was designed to simulate a real scenario (i.e. real SEMG signals do not have gold-standards) but also to test whether inexperienced observers can intuitively learn how to detect events better. The stable levels of high accuracy among blocks and sessions suggest that observers did not extract enough information from signals to significantly improve their accuracy as a function of repetition. Whether the inclusion of feedback information about the observer's performance increases the short- or long-term accuracy of manual segmentation is unknown and it is the focus of ongoing research.

The nearly constant accuracy was accompanied by an improvement in speed for manual segmentation. The group-average learning curve suggested that observers performed manual segmentation progressively faster until a plateau of performance was achieved. The high values of R^2 confirmed that the block number is a good predictor for the learning percentage. This improvement in speed is usually attributed to familiarization with the procedure, best use of equipment, or to the discovery of 'shortcuts' to complete the procedure (21). Based on this study, it is recommended that inexperienced observers train manual segmentation with the same software for at least 300 signals (equivalent to 1 session in this study) before they perform at the fastest rate with neither gain nor loss of accuracy. But most importantly, the lack or training in manual segmentation does not affect the accuracy of observers.

Conclusion

Inexperienced observers exhibit high, not trainable accuracy and a practice-dependent shortening in the time spent to detect the quantity of contraction events in simulated surface electromyograms.

Acknowledgement

References

1. Farina D, Merletti R, Enoka RM. The extraction of neural strategies from the surface EMG. J Appl Physiol. 2004;96(4):1486-95.

2. Merletti R, Parker PA. Electromyography: physiology, engineering, and noninvasive applications. New Jersey: IEEE Press; 2004.

3. Ferreira AS, Guimarães FS, Silva JG. Aspectos metodológicos da eletromiografia de superfície: considerações sobre os sinais e processamentos para estudo da função neuromuscular. Rev Bras Cienc Esporte. 2010;31(2):11-30.

4. Abbink JH, van der Bilt A, van der Glas HW. Detection of onset and termination of muscle activity in surface electromyograms. J Oral Rehabil. 1998;25(5):365-9.

5. Bonato P, D'Alessio T, Knaflitz M. A statistical method for the measurement of muscle activation intervals from surface myoelectric signal during gait. IEEE Trans Biomed Eng. 1998;45(3):287-99.

6. Micera S, Vannozzi G, Sabatini AM, Dario P. Improving detection of muscle activation intervals. IEEE Eng Med Biol Mag. 2001;20(6):38-46.

7. Wilen J, Sisto SA, Kirshblum S. Algorithm for the detection of muscle activation in surface electromyograms during periodic activity. Ann Biomed Eng. 2002;30(1):97-106.

8. Staude G, Wolf W. Objective motor response onset detection in surface myoelectric signals. Med Eng Physics. 1999;21(6-7):449-67.

9. Di Fabio RP. Reliability of computerized surface electromyography for determining the onset of muscle activity. Phys Ther. 1987;67(1):43-8.

10. Marques CK, Terrien J, Rihana S, Germain G. Preterm labour detection by use of a biophysical marker: the uterine electrical activity. BMC Pregnancy Childbirth. 2007;7(Suppl. 1):S5.

11. Moslem B, Khalil M, Marque C, Diab MO. A wavelet package-based energetic approach for the analysis of the uterine EMG. In: ISSNIP Biosignals and Biorobotics Conference; 2010, Vitória; 2010. p. 98-102.

12. De Marchis C, Schimd M, Conforto S. An optimized method for tremor detection and temporal tracking through repeated second order moment calculations on the surface EMG signal. Med Eng Phys. 2012;34(9):1268-77.

13. Shwedyk E, Balasubramanian R, Scott RN. A nonstationary model for the electromyogram. IEEE Trans Biomed Eng. 1977;24(5):417-24.

14. Karlsson S, Yu J, Akay M. Enhancement of spectral analysis of myoelectric signals during static contractions using wavelet methods. IEEE Trans Biomed Eng. 1999;46(6):670-84.

15. Rosa IG, Garcia MAC, Souza MN. A novel electromyographic signal simulator for muscle contraction studies. Comp Methods Programs Biomed. 2008;89(3):269-74.

16. Farina D, Merletti R. Comparison of algorithms for estimation of EMG variables during voluntary isometric contractions. J Electromyogr Kinesiol. 2000;10(5):337-49.

17. Farina D, Fosci M, Merletti R. Motor unit recruitment strategies investigated by surface EMG variables. J Appl Physiol. 2002;92(1):235-47.

18. D'Alessio T, Conforto S. Extraction of the envelope from surface EMG signals. IEEE Trans Eng Med Biol. 2001;20(6):55-61.

19. Delavy DL, Wilson SJ, Richardson CA. Quantification of the timing of continuous modulated muscle activity in a repetitive-movement task. Physiol Measur. 2006;27(11):1143-50.

20. Wright TP. Factors affecting the cost of airplanes. J Aeronautical Sci. 1936;3:122-8.

21. Anzanello MJ, Fogliatto FS. Curvas de aprendizado: estado da arte e perspectivas de pesquisa. Gest Prod. 2007;14(1):109-23.

22. Mewet DT, Reynolds KJ, Nazeran H. Reducing power line interference in digitised electromyogram recordings by spectrum interpolation. Med Biol Eng Comput. 2004;42(4):524-31.

23. Reaz MBI, Hussain MS, Mohd-Yasin F. Techniques of EMG signal analysis: detection, processing, classification and applications. Biol Proced Online. 2006;8(1):11-35.

24. Lu G, Brittain JS, Holland P, Yianni J, Green AL, et al. Removing ECG noise from surface EMG signals using adaptative filtering. Neurosci Lett. 2009;462(1):14-9.

25. De Luca CJ, Gilmore LD, Kuznetsov M, Roy SH. Filtering the surface EMG signal: movement artifact and baseline noise contamination. J Biomech. 2010;43(8):1573-9.

26. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology. 1982;143(1):29-36.

27. Li JJ. A two-step rejection procedure for testing multiple hypotheses. J Stat Plann Infer. 2008;138(6):1521-7.

28. Malone A, Meldrum D, Gleeson J, Bolger C. Reliability of surface electromyography timing parameters in gait in cervical spondylotic myelopathy. J Electromyogr Kinesiol. 2011;21(6):1004-10.

29. Staude G, Flachenecker C, Daumer M, Wolf W. Onset detection in surface electromyographic signals: a systematic comparison of methods. EURASIP J Adv Signal Process. 2001;2:67-81.