Correlation between the stability of feature distribution and classification performance in sEMG signals

Bingbin Wang
Dept. of Engineering
King's Collge London
London, UK
bingbin.wang@kcl.ac.uk

Ernest N. Kamavuako
Dept. of Engineering
King's Collge London
London, UK
ernest.kamavuako@kcl.ac.uk

***Abstract*— The long-term robustness of pattern recognition-based myoelectric systems draws more attention from researchers. Though, there is a lack of analysis investigating how features change over time. This study used two metrics: Coefficient of variation of the first four moments (CoV) and Two-Sample Kolmogorov-Smirnov Test statistics (K-S); to quantify the stability of feature distributions and correlate their changes over time to classification performance. We acquired two surface electromyography (sEMG) channels from sixteen subjects (ten able-bodied and six trans-radial amputees) performing three hand motions. Results showed that the selected metrics correlate to some degree to classification accuracy. Feature distributions are affected less by the time when data are combined. These results imply that stable temporal change may be an acceptable way to choose robust features in long term investigations.**

Keywords—Long-term, Pattern recognition, Surface Electromyography, Feature, Myoelectric control

I. Introduction

Over the past twenty years, pattern recognition based myoelectric prostheses have developed rapidly. Many researchers focus on improving the classification performance of myoelectric-controlled prostheses in acute settings[1]. The classification has achieved over 90% accuracy based on researchers' proposed methods with dramatic performance degradation in clinical applications[2]. For example, Kaufmann et al. (2010) [3] investigated the performance of the state-of-art pattern recognition algorithm on 21 days of sEMG data. Results showed that classification accuracy dropped gradually with the initially trained classifier. Several causes of long-term sEMG variation have been explored, such as electrodes shift [4], muscle fatigue [5], arm position [6], user adaptation [7] etc.

In the pattern recognition-based scheme, feature extraction is crucial to extract discriminable information. The comparison of utilising different classifiers and feature selection illustrated that feature selection considerably influences classification performance [8]. Variations between training data and testing data will lead to inconsistencies in feature space, which will directly interfere with the classifier's estimation of different classes of signals. In a past study, Phinyomark et al. (2013) [9] evaluated 50 time-domain and frequency-domain features using sEMG recorded during 21 days to determine the best robust single feature and feature set. Results showed sample entropy and proposed feature set could achieve 93.37% and 98.87% classification accuracy, respectively, with only initial first-day training. In addition, Tkach et al. (2010) [8] simulated the physical and physiological changes that could occur in daily use and investigated the stability of time-domain features against these changes. In their thorough analysis, they demonstrated the impact of these changes on classification accuracy. Sequentially, based on these results, the most stable feature and feature set are also selected.

Despite these recent findings of the role of stability of long-term classification performance, the research outputs still cannot meet the clinical requirement. The neglected aspect of long-term sEMG study is how features changes can predict performance. Quantifying feature characteristics can effectively assist people to understand feature space variation in the context of long-term applications. In this paper, we aim to assess whether the stability of feature distributions correlate with performance. The analysis was based on the pseudo-true distribution functions constructed using kernel estimation. The stability of two distributions is quantified using the following metrics: Coefficient of variation of the first four moments (CoV) and Two-Sample Kolmogorov-Smirnov Test statistics (K-S).

II. Methods

A. Subjects

This analysis used previously reported data by Waris et al. (2018) [10]. Ten able-bodied subjects (all male) and six trans-radial amputees (all males, three left and three right-hand amputations) participated.

B. Data Collection

Surface EMG signals were acquired by a commercial acquisition system (AnEMG12, OT Bioellectronica, Torino, Italy), filtered (10-500Hz) and sampled at 8000 Hz. The entire experiment consisted of seven sessions throughout a week (seven days). Subjects performed 11 motions repeatedly up to four times in each session. Each hand motion was sustained for up to five seconds with five seconds rest in-between movements. This study only uses two channels (one flexor and one extensor) and three classes

Bingbin Wang was sponsored by the China Scholarship Council and King's College London.

(Rest, Open Hand, Close Hand) as a feasibility investigation.

C. Signal Processing and Feature Extraction

Surface EMG signals were digitally high-pass filtered (fourth-order Butterworth filtered) between 20 and 500 Hz. We reduced power line interferences using a notch filter at 50 Hz. From every five seconds of contraction time, the first second was designated as the onset phase and the last second as the offset phase to avoid non-stationarity. Subsequently, three seconds of the steady-state phase per repetition were used for the extraction of features. Forty features were extracted from incrementing (by 40 ms) windows of 200 ms duration. These features were: Integral Absolute Value, Mean absolute value (MAV), Modified Mean Absolute Value type 1 (MAV1), Modified Mean Absolute Value type 2 (MAV2), Simple Square integral (SSI), Variance (VAR), Absolute Value of 4th Temporal Moment (TM4), Root Mean Square (RMS), v-Order(3rd) (V3), Log Detector (LOG), Waveform Length (WL), Average Amplitude Change (AAC), Difference Absolute Standard Deviation Value (DASDV), Amplitude of the first burst (AFB), Zero Crossing (ZC), Myo-pulse Percentage Rate (MYOP), Willison Amplitude (WAMP), Slope Sign Change (SSC), Cepstral coefficients (CC), Approximate entropy (ApEn), Detrended fluctuation analysis (DFA), Higuchi's fractal dimension (HFD), Katz's fractal dimension (KFD), Kurtosis (KT), Maximum fractal length (MFL), Sample entropy (SpEn), Skewness (Skew), Maximum amplitude (MA), Mean Frequency (MNF), Median Frequency (MDF), Peak Frequency (PKF), Mean power (MP), Total power (TP), The 2nd spectral moments (SM2), Frequency ratio (FR), Power spectrum ratio (PSR), Variance of central frequency (VCF), Maximum to minimum drop in power density ratio (MMDP), Power spectrum deformation (PD), Signal-to-noise ratio (SNR).

The extracted features are mainly in the time domain and the frequency domain. Since the time-frequency domain features are sensitive to fluctuations and result in low classification performance [11], it was not considered in this research.

D. Distribution functions and data analysis

Data analysis was performed on a single feature basis. Data were divided into training sets composed of incremental cumulation of days from Day1 to Day6, and the test set contained Day7 only, as depicted in Fig 1. This study investigated whether a shorter interval between training and testing set would result in better classification performance. Hence, we chose to add new days to the training set instead of taking random combinations between days. For each set (training or testing), the distribution function (or probability density function, PDF) was constructed using kernel estimation. The first four moments of each distribution were then computed for each distribution. To compare the distribution of training set with the distribution of the test set, the following metrics were computed:

(1) Coefficient of Variation (CoV): the CoV of moments between the PDF of the training set and PDF of the testing set was computed. The CoV of all moments were pulled together and averaged.

(2) Two-sample Kolmogorov-Smirnov test statistics (K-S) were used to access the similarity between the PDFs.

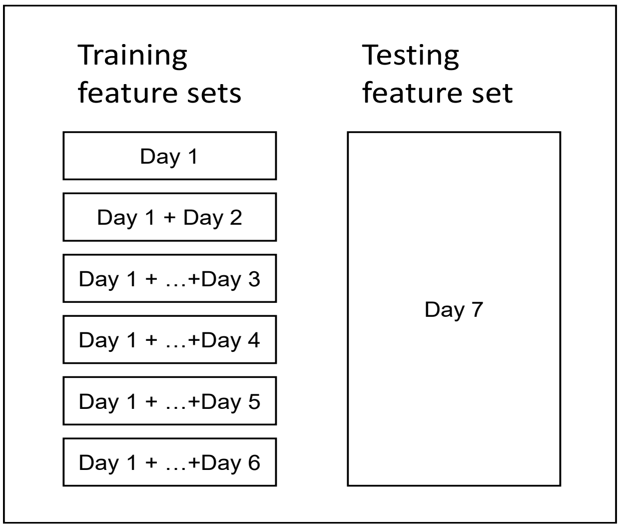

Fig. 1. The pipeline of feature distribution analysis based on cumulative training set and a single test set (Day7). There were six sets of results for each analysis corresponding to different training feature sets.

Each metric is an average over channels, motions and days to account for the multi-dimension of the feature space. We chose the cumulative training set as it shows stable distributions, as shown in Fig 2.

Linear discriminant analysis (LDA) was used as a classifier due to its simplicity and robustness compared to other

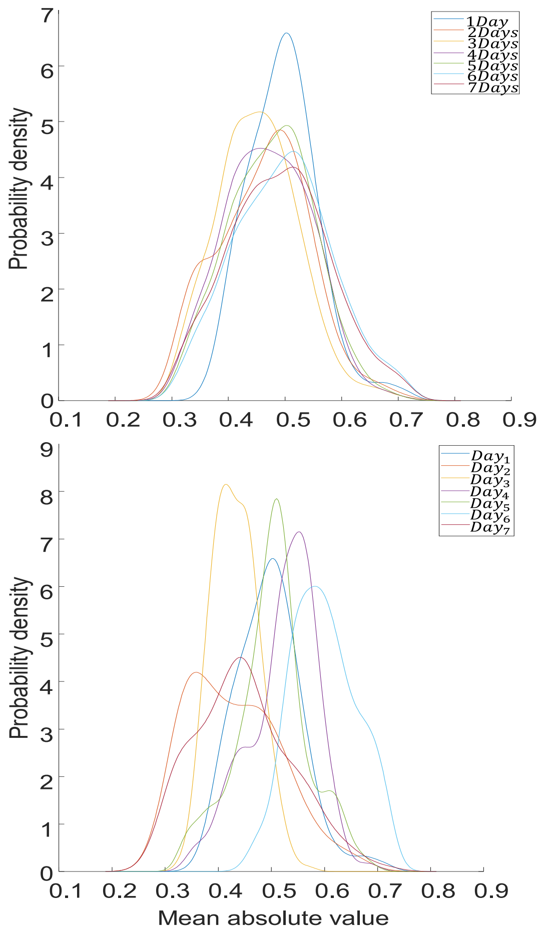

Fig. 2. Probability density functions of mean absolute value, the above one is cumulative days distribution and below one is single days

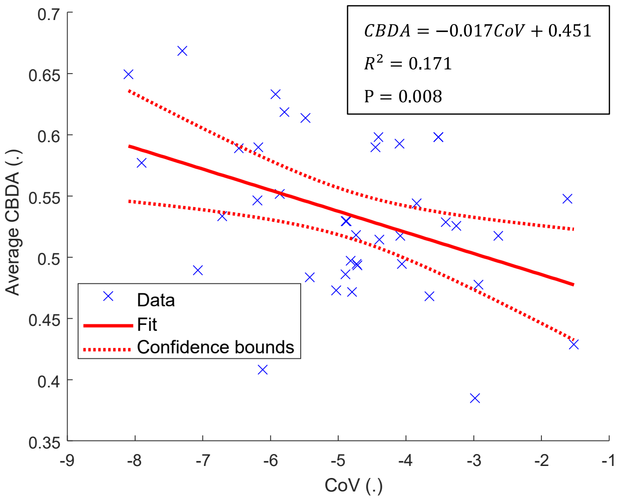

Fig. 3. Scatter plot of a linear regression model where the Combined Days Accuracy (CBDA) is related to the log of Coefficient of Variance (CoV) of the first four moments

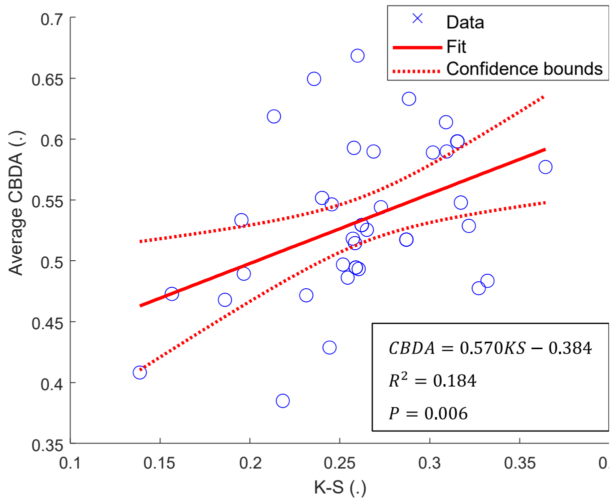

Fig. 4. Scatter plot of a linear regression model where the Combined Days Accuracy (CBDA) is related to the Kolmogorov-Smirnov test statistics (K-S)

conventional classifiers [3][12]. Between-days classification was carried out to investigate the effect of time on classification performance. In between-day classification, the classifier was trained by the cumulative days training set and tested with the seventh day's data, as Fig.1 shows.

We fitted a line between the average values over all subjects of each metric as independent variables against the average cumulative between-days accuracy (CBDA). R-squared and p-values were estimated for each linear model. Subsequently, multiplying inverse CoV and K-S values with CBDA, respectively, the ten largest features were selected as the robust features for each metric.

III. Results

A. *Probability density function*

Probability density functions of all forty features were calculated and plotted for each movement, channel and subject. Fig. 2 shows an example of probability density plots of mean absolute value on a single day and cumulative days datasets. We note that the PDF changes over days, but can be kept stable when combining days.

B. *The first four central moments & Kolmogorov-Smirnov test statistics*

For CoV and K-S (Fig. 3, Fig. 4), a proportional effect has been observed where features with stable distributions (lower values of CoV and higher values of K-S) tends to perform better than features with higher values of CoV and lower values of K-S. The reported R^2 is 0.171 for CoV and 0.184 for K-S, respectively. Their values indicate a very low degree of linearity, but the slopes are significant ($P < 0.05$).

In addition, The ten features with the best correlations are presented in Table I. Features that appear twice in the two metrics are Katz's fractal dimension (KFD), Approximate entropy (ApEn) and Detrended fluctuation analysis (DFA).

IV. Discussion

Extensive studies investigated the longevity of EMG signals and tried to mitigate the effect induced by time on classification performance. Researchers tested and used various state-of-the-art classifiers in multi-day classification [3][9], evaluated the long-term performance of different types of EMG signals[12], developed novel deep learning techniques to improve classification performance over time [13][14]. All of them used classification accuracy as criteria to demonstrate the time-induced effect in EMG signals. However, An essential element missing from these studies is the change in feature space over time. Classification accuracy depends on the discriminant ability of classifiers and the quality of features. Hence, it will be more intuitive to inspect changes in long-term EMG signals from the perspective of feature space. He et al. (2015) [15] used

TABLE I. The ten best robust features for CoV and K-S and corresponding CBDA, CoV and K-S values

Metrics \ Rank	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th
CoV	HFD	**KFD**	**DFA**	MYOP	PD	**ApEn**	SpEn	SSC	MNF	ZC
CoV value	0.0003	**0.0004**	**0.0007**	0.0008	0.0012	**0.0016**	0.0021	0.0020	0.0027	0.0030
CBDA	0.6495	**0.5772**	**0.6685**	0.4894	0.5335	**0.5890**	0.5899	0.5464	0.6332	0.6186
K-S	**KFD**	MDF	MFL	AAC	WL	MNF	DASDV	**ApEn**	MMDP	**DFA**
K-S value	**0.3644**	0.3091	0.3153	0.3153	0.3153	0.2884	0.3094	**0.3017**	0.3174	**0.2598**
CBDA	**0.5772**	0.6139	0.5981	0.5981	0.5981	0.6332	0.5898	**0.5890**	0.5478	**0.6685**

separability index and repeatability index to quantify the feature space changes and correlated them with multi-day classification performances. They demonstrated that decreased variations in EMG feature space resulted in gradually increased between-day classification performance. As their primary focus was on user adaptation, the degree of linearity between the two indices and multi-day performance was not illustrated. This study achieved quantification of feature change for long-term EMG investigation. The probability density functions visualised the change in EMG over multiple days. For easy interaction, the first four central moments are quantitative measures to describe each feature distribution. Kolmogorov-Smirnov test statistics quantified variations between two distributions. Both metrics exhibited some degree of correlation with performance.

However, these results were low compared to Phinyomark's study[9]. The low accuracy might be due to only two channels of sEMG signals and the average over multiple scenarios. Hence, it is necessary to discard low-quality data from the training set to prevent performance degradation. This investigation is limited to using one-dimensional metrics that are averaged to account for multidimensionality. More research is needed to determine intuitive metrics to quantify multidimensional probability distributions.

V. Conclusion

In this study, we presented an investigation of long-term classification performance. The paper demonstrated the potential of the first four moments and Kolmogorov-Smirnov test statistics to correlate with performance. In conclusion, stability of feature distribution can be a way to predict performance behaviour, not just with time but to other factors such as limb position or fault detection in general.

References

[1] Simao, M., Mendes, N., Gibaru, O. and Neto, P., 2019. A Review on Electromyography Decoding and Pattern Recognition for Human-Machine Interaction. IEEE Access, 7, pp.39564-39582.

[2] Campbell, E., Phinyomark, A. and Scheme, E., 2020. Current Trends and Confounding Factors in Myoelectric Control: Limb Position and Contraction Intensity. Sensors, 20(6), p.1613..

[3] Kaufmann, P., Englehart, K. and Platzner, M., 2010. Fluctuating emg signals: Investigating long-term effects of pattern matching algorithms. 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology.

[4] He, J., Sheng, X., Zhu, X. and Jiang, N., 2020. Position Identification for Robust Myoelectric Control Against Electrode Shift. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 28(12), pp.3121-3128.

[5] Díaz-Amador, R. and Mendoza-Reyes, M., 2019. Towards the reduction of the effects of muscle fatigue on myoelectric control of upper limb prostheses. DYNA, 86(208), pp.110-116.

[6] Betthauser, J., Hunt, C., Osborn, L., Masters, M., Levay, G., Kaliki, R. and Thakor, N., 2018. Limb Position Tolerant Pattern Recognition for Myoelectric Prosthesis Control with Adaptive Sparse Representations From Extreme Learning. IEEE Transactions on Biomedical Engineering, 65(4), pp.770-778..

[7] Hahne, J., Markovic, M. and Farina, D., 2017. User adaptation in Myoelectric Man-Machine Interfaces. Scientific Reports, 7(1).

[8] Tkach, D., Huang, H. and Kuiken, T., 2010. Study of stability of time-domain features for electromyographic pattern recognition. Journal of NeuroEngineering and Rehabilitation, 7(1), p.21.

[9] Phinyomark, A., Quaine, F., Charbonnier, S., Serviere, C., Tarpin-Bernard, F. and Laurillau, Y., 2013. EMG feature evaluation for improving myoelectric pattern recognition robustness. Expert Systems with Applications, 40(12), pp.4832-4840.

[10] A. Waris et al., 2018. "The effect of time on EMG classification of hand motions in able-bodied and transradial amputees", Journal of Electromyography and Kinesiology, vol. 40, pp. 72-80. Available: 10.1016/j.jelekin.2018.04.004.

[11] Phinyomark, A., Phukpattaranont, P. and Limsakul, C., 2012. investigating long-term effects of feature extraction methods for continuous emg pattern classification. Fluctuation and Noise Letters, 11(04), p.1250028.

[12] A. Waris, I. Niazi, M. Jamil, K. Englehart, W. Jensen and E. Kamavuako, 2019. "Multiday Evaluation of Techniques for EMG-Based Classification of Hand Motions", IEEE Journal of Biomedical and Health Informatics, vol. 23, no. 4, pp. 1526-1534. Available: 10.1109/jbhi.2018.2864335.

[13] Zhai, X., Jelfs, B., Chan, R. and Tin, C., 2017. Self-Recalibrating Surface EMG Pattern Recognition for Neuroprosthesis Control Based on Convolutional Neural Network. Frontiers in Neuroscience, 11.

[14] U. Cote-Allard et al., 2019. "Deep Learning for Electromyographic Hand Gesture Signal Classification Using Transfer Learning", IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 27, no. 4, pp. 760-771. Available: 10.1109/tnsre.2019.2896269.

[15] He, J., Zhang, D., Jiang, N., Sheng, X., Farina, D. and Zhu, X., 2015. User adaptation in long-term, open-loop myoelectric training: implications for EMG pattern recognition in prosthesis control. Journal of Neural Engineering, 12(4), p.046005.