Towards inter-session hand-gesture classification

A review on the state-of-the-art of the challenges and technologies using EMG

David Yenicelik
ETH Zurich
yedavid@ethz.ch

ABSTRACT

This review is an extension to the survey conducted by Adrian Spurr in 2015 [**?**] on using electromyography (EMG) signals in machine learning algorithms with a focus on HCI and hand pose estimation. This paper gives a broad overview of what has happened since 2015, but will partially overlap with contents presented in Spurr's survey. As such, the following is a copy of the abstract of Spurr's paper: "This paper serves as a survey for using electromyography (EMG) signals in machine learning algorithms, with a focus on HCI and hand pose estimation. It gives an overview on the current applications and issues which have been tackled, lists problems occurring when using this type of input and names a series of open questions which have not been sufficiently addressed or solved in a satisfiability manner. It ends with a proposal for a novel feature for approximating continuous hand pose with EMG, motivated by the current literature available..."

Keywords: surface Electromyography, high-density EMG, cross-session, hand-gesture classification, active prosthetics, HCI

PROPERTIES OF sEMG SIGNALS

Electromyography (EMG) sensors record electrical signals formed by muscles that originate from the peripheral nervous-system. Raw sEMG signals have peak-to-peak amplitudes of 0-2mV, rising up until 10mV [**?**] with a band frequency ranging between 0-1000Hz [**?**]. The band frequency of the signal which includes significant information has a range of 20-500Hz [**?**] [**?**] or 50-150 Hz [**?**]. sEMG signals are time and subject varying because of muscle fatigue, muscle length, muscle fibre architecture and level of effort [**?**] [**?**], as well as electrode shifts, changes in arm-posture and slow time-dependent changes such as electrode-skin impedance [**?**] [**?**]. The normalized sEMG amplitude is often obtained through non-linear division by the respective Maximum Voluntary Contraction (MVC) of the test subject [**?**] [**?**]. The MU firing rate follows a non-linear law [**?**] and the number of MUs activated is determined by the power of the muscle contraction [**?**]. There is an uncertain relation between EMG signals and muscle application [**?**]. sEMG signals are time dependant [**?**] and due to the mentioned reasons above believed to be stochastic [**?**] [**?**][**?**] [**?**].

NOISE SIGNALS RESULTING FROM EMG MEASUREMENT (NAZMI et al. 2016)

There are two main issues that influence the fidelity of the signals. Signal-to-noise ration examines the ratio of energy in EMG signals to energy in noise signals. Yet another problem are pure noise signals, of which there are different categories:

Inherent Noise in Electronics Equipment in EMG Signals is caused by electrical equipment. Electrodes made of silver (Ag) and silver chloride (AgCl) however are found to give an adequate signal-to-noise-ratio (SNR) and are electrically very steady. Furthermore, bigger electrode sizes imply decreased impedance. Generally, this type of noise can be eliminated using intelligence circuit design and high quality instruments [**?**].

Ambient Noise is 1 to 3 times bigger than the EMG signal of interest, which includes Power-Line Interference (PLI) arising from the 50-60Hz spectrum. A high pass filter can remove the interference if the frequency of the interference is high [quantify high!] [**?**]. This is the most common kind of noise [**?**].

Motion Artifacts are caused by moving muscles, skin and spread of the innervation zone (IZ) electrode displacement [**?**] [**?**] [**?**]. Blood flow can also be a factor [**?**]. These can be reduced by proper design and setup of the system. For example, band pass filters with a high-pass filter having a 500Hz cut-off frequency and a low pass filter having a 20Hz cut-off frequency already reduce these [**?**] [**?**].

Inherent instability of the signal is caused by the stochastic nature of the EMG signals [**?**]. It can be regarded as unwanted because this signal unstable, usually lying within the range of 0-20 Hz [**?**], but can potentially be accepted without a de-noising structure within the system.

Electrocardiographic (ECG) Artifacts arise due to an overlap of ECG and EMG frequency spectra and their shared characteristics such as non-stationarity and varying temporal shape [**?**] [**?**]. It is very difficult to remove ECG [**?**]. However, if ECG can be effectively measured, this signal could be subtracted from the EMG signal. Bandpass filters and mathematical morphology operator (MMO) appear to bring acceptable results in filtering [**?**] [**?**].

Cross-Talk arises when signals from nearby muscles, which are not in the monitored muscle group, are recorded by the

electrode [**?**]. Setting the IED to 1-2cm, and carefully adjusting the or the radius of the electrode can reduce this noise effectively [**?**].

There is no generally applicable de-noising strategy. As such, de-noising is application-specific.

MATHEMATICAL MODEL OF MEASURED EMG SIGNALS

There are hypothesis for different mathematical models that describe EMG signals. However, simulating models lack realism, especially when simulating MU firing rates, as well as force generation phases [**?**]. Some refer to probability density functions that can describe the amplitude of EMG signals [**?**] [**?**]. The classification accuracy increases when the signal is measured between 128-500ms [**?**][**?**] [**?**]. The resulting model is related to the electrode positioning [**?**], so it's not always possible to solely talk about the nature of EMG signals alone, but the measured EMG signals. Although the EMG signals are believed to have a reproducible [**?**] stochastic nature [**?**] [**?**] [**?**], the investigated literature does not show consensus among the chosen Probability Density Function (PDF) of the signals [**?**]. Although the central limit theorem proves that any dataset can be modelled as a normal distribution, it has been shown that the distribution of EMG generating signals is not a Gaussian one. This is because the actual signals is skewed towards zero [**?**] [**?**].

PARAMETERS FOR ELECTRODES (Kilby, Prasad, and Mawston 2016)

The following paragraphs enumerate and discuss common variables in electrode-placement.

Electrode Design and Configuration Electrode materials that are used most often include pre-gelled silver (Ag) or silver chloride (AgCl). For dry configurations, no preference is detected [**?**], but similar performance to gelled electrodes are shown [**?**]. Shapes are of minor significance. SENIAM has recommendations on the setup of cables, electrodes and a possible pre-amplifier [**?**]. Increasing the number of electrodes increases accuracy, but plateaus quickly after 100 placed electrodes [**?**].

Data Acquisition of sEMG signals include sampling frequency and bit-resolution of the data acquisition card [**?**].

The choice of monopolar, bipolar, single-differential or double-differential is another parameter to be set, especially in the case of grid-like structures. Usually an electron put on a bony area is used as a ground-signal. As such, studies are also conducted on the placement of the electrodes relative to the innervation zones (IZ) and tendon zones (TZ) [**?**] [**?**]. Monopolar detection provides the maximum information, but also includes fiber end effects, while bipolar detection provides a clear picture of any innervation and tendon zones. However, double differential detection is the most suitable for estimating the muscle fiber conduction velocity [**?**].

Signal Pre-Processing of sEMG signals [**?**] can vary from simple bandpass filters [add reference here], butterworth filters [add reference here] to signal amplifications [add reference].

Factors in the testing of the subject include the muscle-specific deviations, the experimental set-up, and the placement of the electrodes on the skin surface [**?**]. Sometimes, small muscles might have a high impact on the gesture formed, but might not produce high-enough threshold levels [**?**]. Possibly, this is caused because these muscle are too deeply located [**?**].

Generally, the procedure of placing the electrodes on the skin is a topic that is not much discussed in current literature [**?**] [**?**]. Again, these factors are algorithm- and application-specific. See appendix A for a list of manufacturers.

COMMON FEATURES USED (Nazmi et al. 2016)

Features can be divided into time-domain (TD), frequency-domain (FD) or time-frequency-domain (TFD) (Tsai et al. 2014) (Hogan and Mann 1980) (Englehart, Hudgins, and Parker 1999) (Nazmi et al. 2016). The following will discuss the most often used feature, the reader can refer to (Nazmi et al. 2016) for a table of features used in previous papers. None of the features selected seem to be an automated choice (Nazmi et al. 2016).

Root mean squared (RMS) (Kendell and Lemaire 2012) (Nazmi et al. 2016) is a TD-feature that is often used with bipolar single EMGs when a relation between force and Muscle Unit (MU) is investigated upon [cite all the papers that do this]. For sEMG grids, experiments often average over multiple time frames of sEMG frames to receive one instantaneous image that represents this timeframe in one image. ?RMS appears to be the best parameters compared to the others, as it provides a quantitative measure for electrode selection? (Nazmi et al. 2016). It is often used to measure raw muscle activation, without specification of the MUs or locations of the muscles [add some references here].

Mean Frequency (MNF) and Median Power Frequency (MNP) are FD-features and are often used to characterize EMG signals, especially for muscle contractions (Merletti and Conte 1997), or to characterize fatigue over time (Phinyomark and Limsakul 2009). However, different MUs have different recruitment thresholds, and thus must be considered differently (Linnamo 2002) (Kossev and Christova 1998) (Nazmi et al. 2016).

Time-Frequency-Domain (TFD)-features can either be novel features or ensemble techniques of the above explained TD and FD features (Guo and Kareem 2016). However, these features suffer from the curse of dimensionality (Boccia et al. 2015), which can be reduced through dimensionality reduction techniques (Nazmi et al. 2016) or potentially an auto encoder finding an optimal embedding layer. [add this to the discussion rather maybe?]

GRID-CONFIGURATIONS OF sEMG SIGNALS (Kilby, Prasad, and Mawston 2016)

There are different types of grid-configurations for array-shaped sEMGs. A combination of multiple, for example 8 linear arrays put around the forearm is also possible [add reference here]. Sizes in arrays usually differ greatly but are between 1-10mm, either circular or elliptical in shape. SENIAM has no suggestions for the inter-electrode-distance

(IED) sEMG?s in grids yet, but past experiments show that values between 2.5-20mm are common (Kilby, Prasad, and Mawston 2016).

Linear Array Electrodes can use monopolar, bipolar and single (or more) differential configuration. Often stacked together columns wise into an arrays.

2D Array Electrodes usually produce monopolar signals which are then further processed for analysis (Kilby, Prasad, and Mawston 2016).

HD-sEMG are generally electrodes that are uniform in two spatial axes and densely distributed. 2D multi-electrode configuration exhibit higher signals and lower cross-talk compared with other types (Dimitrov, Disselhorst-Klug, and Dimitrova 2003). An intended consequence of bipolar and higher-order montages, and of a short IED, is the suppression of the far-field activity (Kilby, Prasad, and Mawston 2016).

HSR-sEMG electrodes usually record monopolar signals and process these through convolutional weight matrices (usually 4 for centre and -1 for neighbouring pixels). This is equivalent to the Laplacian operator and approximates second-(or higher)-order differentials. Applied along the fibre direction, this can also approximate conduction velocity, and determine the location of innervation zones and tendons.

COMMON PRE-PROCESSING METHODS

Most studies don?t put much detail into the pre-processing stage, but it has repeatedly been shown that Butterworth filters with a cut-off frequency of 20-500Hz has been used (Boxtel 2001) (Kim, Kim, and Kim 2016). However, this ?noise? might also include useful patterns (Geng et al. 2016), and should be removed with care. Alternatives are bandpass filter with frequencies of 5-322Hz (Sulaiman et al. 2016) or similar (20-380Hz (Du et al. 2017)). Power-Line interference might be remove using a band-stop filter (45-55Hz second order Butterworth) (Du et al. 2017). The signal might also need to be converted from analogue to digital through a ADC converted (Sulaiman et al. 2016). Segmentation techniques are relative to the application. It is open to discussion, if adjacent or overlapped windowing techniques are preferable (Nazmi et al. 2016). As always, finding the optimal features is also an open question, but there exists a preference.

HAND CLASSIFICATION MEHOTDS USING EMG SIGNALS

Some major work has been done in the field of within/intra-session (test and training-data taken from within one session and one subject), hand gesture classification. The majority of the work has been done in classification, while there also exists work in continuous estimation (El-Khoury et al. 2015). Some common algorithms that have shown good results include:

- **Latent Dirichlet Algorithm (LDA)** is being deployed after using PCA to reduce to input feature set with a classification accuracy of 93.75% for [number of gestures] different gestures [cite actual papers] (Nazmi et al. 2016).
- **Multi-Layer-Perceptron (MLP)** is being used achieving 99% accuracy for 6 different hand gestures to be classified (Khushaba and Al-Jumaily 2007) (Nazmi et al. 2016).
- **Fuzzy Logic (FL)** classifies the stages of contraction (start, middle, end) with 97% accuracy [cite actual papers] [cite number of gesture] (Nazmi et al. 2016). Neuro-Fuzzy systems are an extension to this (Khezri and Jahed 2011).
- **Ensemble techniques** using multiple feature sets even achieve 98.87% classification accuracy [look how many different hand gestures were included] [cite actual papers] (Nazmi et al. 2016).
- **Support Vector Machines (SVM)** is being deployed to classify instantaneous frames from sEMG measurement (Saponas et al. 2010), or are sometimes ensemble with pressure sensors (FSR) to also include wrist (McIntosh et al. 2016) or forearm [add reference with IMU] motion. For included pressure sensors accuracies range between 97.5% for wrist gestures and 94.0% for finger gestures for 14 gestures. For inertia sensors accuracies range between [must look this up again, should be in the IMU paper].
- **Deep Convolutional Neural Nets** can be used in ensemble over 40 frames at 1000Hz, using simple majority voting to get classification accuracies of over 96.8% for over 40 gestures (Geng et al. 2016). Apparently, 150ms is the windows size suggested to consider, to classify a gesture (Geng et al. 2016). All this is possible because the HD-sEMG can be considered as an imaging tool (Merletti, Holobar, and Farina 2008). It is also reported that usually we have 8-11 firing MUs per second (Martinez-Valdes et al. 2016) which might satisfy why recordings of 100ms is needed.

Funnily, some algorithms, such as the Fuzzy Logic system, improve when the number of output classes are increased (El-Khoury et al. 2015)

APPLICATIONS OF INTER/CROSS-SESSION sEMG

Also, motions are sometimes detected through ultrasound before sEMG signals take place, however, this could also be a bug in the calibration, due to the algorithms, or because initial bursts were ignored (Dieterich et al. 2017). Some extensive studies have been done, correlating with ultrasound sensors, showing that different time-windows should be considered when classifying an EMG signal (Dieterich et al. 2017).

ACTUAL APPLICATIONS OF INTER/CROSS-SESSION sEMG

We will present challenges that must be overcome to solve the cross-session problem, and show past approaches that tried to solve this problem. Although (auto-)calibration features are desirable, [add reference here, Amma, Microsoft, Kernel guys, etc.]. little work has been done on cross-session sEMG measurements, so (Martinez-Valdes et al. 2016) to find papers that included cross-session application, papers that included the ”term”, ”days” or ”weeks” were taken into consideration.

Challenges

Often, different subjects have different muscle structures with which some muscles are more visible to the device than others. (Martinez-Valdes et al. 2016). A benchmarking dataset is missing, and the biggest one contains 23 participants with [number of gestures] gestures each (Du et al. 2017). The common problem of inter-subject, and even in cross-session, sEMG measurements is the high subject-specifity and high variability of the attained data (Castellini and Van Der Smagt

2009) (Farina, Jiang, and Rehbaum 2014). A negative correlation between the body-mass-index and gesture classification accuracy is noted (Atzori, Gijsberts, and Kuzborskij 2015) (Holobar, Minetto, and Farina 2014).

Approaches

Often, some kind of initial calibration (Saponas et al. 2010) or a reference to a relaxed state (Dieterich et al. 2017) is used in cross-session settings . (Amma et al. 2015) pioneered auto-calibration, and a first end-to-end model has been created (Du et al. 2017), but the problem of cross-session hand gesture classification is still an open one. Both presented models recorded sEMG frames over time, and the appropriate output label. A list of datasets can be obtained in Appendix B.

(Amma et al. 2015) tried to measure sEMG in cross-session using a single calibration gesture. For this, he used a bony-area as a reference to shift the resulting sEMG image for calibration. The recognition of the bony-area is only approximated by the area of lowest muscle activity, and found through the watershed algorithm. Another method he used was a Gaussian Mixture Model (GMM) in 2 dimensions, which would identify areas of high activity, and shift the sEMG image accordingly. RMS was used to detect muscle activity, and the data is interpolated using bicubic interpolation. These methods resulted a calibration method with 75% accuracy in hand gesture classification.

(Du et al. 2017) trained a deep Convolutional Neural Net (CNN) using domain adaption (Patel, Gopalan, and Li 2015) techniques. These techniques rely of fine-tuning pre-trained networks (Donahue et al. 2014) . An algorithm that is being used is AdaBN, making the network more robust to inconsistencies in the covariance of the covariance data. AdaBN is similar to batch-normalization, which subtracts the mean and divides the standard-deviation from the incoming layer [but accuracy in here]

$$v = \frac{(u-\mu)}{\sigma} \times \gamma + \beta$$

with the parameters and to be learned. This is regraded as an unsupervised method which does not require labeled sEMG data. During training, it must be ensured that the entire batch stems from the same session and same subject, otherwise, the learning parameters gamma and beta cannot be effectively learned. This model achieved state-of-the-art in cross-session with about 80% accuracy. The auto-calibration takes about 10 seconds to take place.

Other work has also been done, but was less successful (Patricia, Tommasit, and Caputo 2014) (Ju, Kaelbling, and Singer 2000)(Khushaba 2014), posing interesting ideas to consider however that mostly regard domain adaption as an option. Data augmentation (Hargrove, Englehart, and Hudgins 2008) (Boschmann and Platzner 2012) and model adaption techniques (Amma et al. 2015) (Patricia, Tommasit, and Caputo 2014) (Ju, Kaelbling, and Singer 2000)(Khushaba 2014) have been devised for this. Due to this high subjectivity, the problem of cross-session hand gesture classification can be regarded as a multi-source domain adaption problem (Patel, Gopalan, and Li 2015).

DISCUSSION

To produce a wristband that can accurately classify hand gestures, a predictive model needs to be. The problem of creating this pre-trained model can be regarded as a supervised learning task. This pre-trained model can then be used as a starting point, from which calibration for individual subjects can automatically take place, potentially through unsupervised learning.

Future fields of research

Further research could be conducted in the following domains:

Image and Video classification: is a viable path, for the reason that the sEMG signal produced is a 2D matrix of one-dimensional color-values, and as such, resembling an image. Video classification is viable as well, because the sEMG signal produced is time-dependent, a sequence in time, and due to the stochastic nature of the signal, multiple consecutive frames must be drawn to attention to draw a conclusion.

Domain Adaption is a possible research field, as it is believed that changing sessions / subjects just shifts the classification task in space, but that the classification task stays the same. If enough different domains can be covered in the training data, potentially domain adaption techniques could give the small necessary push needed to realize cross-session hand-gesture classificaiton.

Zero-Shot-Learning is a field close to domain adaption, but deals with minor or no labeled data. The goal is to achieve a high-enough learning effect given sparse data to draw personalized quality. This is interesting, because it is possible to get unlabeled data from the user: If it is possible to use this unlabeled data to improve the quality of the calibration, it might be possible to generalize between sessions, and potentially subjects.

Methods for hand gesture classification are interesting because some methods might generalize to cross-session or cross-subject settings.

Possible next steps

Increasing the dataset size: It is often mentioned that increasing the number of data samples fed to a model will increase it's accuracy, assuming the model has a high-enough capacity. As such, increasing the number of samples collected for a given gesture should result in an increased accuracy. Apart from that, it is assumed that there exists a finite number of different muscle types that the model has to adapt to. The dataset with the highest number of different subjects is the NinaPro dataset with about 30 subjects. However, as we can see from other deep architectures, this is usually a challenging small amount of data to rely good learning results for a deep architecture. As such, it would make sense to create a dataset that includes more different sessions (100), and then potentially expand to having a higher number of different test subjects.

Have non-mobile/ classification measure: If it is possible to set up an effective classification measure which can be used naturally for many hours, much more training data can be

collected. However, it should be noted that the current training data should vary in the number of subjects involved, not necessarily number of gestures occurring. Also, any other armband-internal system that collects data, or concurrently improves the classification accuracy of the armband can help. (such as the 2-segment armband, where segment 2 helps classify it, and the neural net is actually trained/backpropagated on segment 1)

Models that could potentially work in classifying cross-session

CNN with LSTMs or Residual/Highway Layers: One could use CNN?s to bring condense the retained momentary image into a hidden feature vector that contains the condensed information about the image. This hidden feature vector can then be passed through multiple LSTMs or through a residual/highway layer for sequence analysis. [Maybe explain more about why this could be a good model] Time is probably an important factor to regard, so probably pay more attention here

A deep architecture within an adaption algorithm: Reinforcement learning uses deep models to approximate the state-space and incorporates the results tabular algorithms. A deep model could also be used in this case to classify a gesture, and proven model adaption algorithms could be used to ?translate? or ?calibrate? the gestures. [Maybe explain more about why this could be a good model]

Wide and Deep Nets: Recommendation systems that generalize but also personalize have a deep and wide architecture that acts similar to an ensemble, except that backpropagation starts from the same end (one unified model). These kinds of models could potentially be useful, however, it could be difficult to implement these effective together with the CNN. Any other architecture that includes a part that personalized, and a part that generalizes could be optimal. Generally, looking at how other domains solve the problem of auto-calibration might be useful. Maybe explain more about why this could be a good model]

Appendix A: Manufacturers list

- Electrodes, mainly for linear arrays: LISiN-Specs Medica, Italy (Kilby, Prasad, and Mawston 2016)
- ActiveOne, BioSemi, Amsterdam, Netherlands (Kilby, Prasad, and Mawston 2016)
- Shinystone, LogOnU Inc. (Kim, Kim, and Kim 2016)
- Inertial Unit: IMU EBIMU24GV2, E2BOX Co. (Kim, Kim, and Kim 2016)
- Wireless SHIMMMER EMG (Sulaiman et al. 2016)
- SPES Medica, Salerno, Italy (Martinez-Valdes et al. 2016)
- ADC: EMG-USB 2, 256-channel EMG amplifier, OT Bioelettronica, Torino, Italy (Martinez-Valdes et al. 2016)
- Delsys Trigno Wireless EMG (El-Khoury et al. 2015)
- OT- Bioelettronica (Amma et al. 2015)

Appendix B: Available datasets that include cross-session data

- NinaPro - Sparse hand prosthetics dataset (Du et al. 2017)
- CSL-HDEMG - Near-to-realistic dataset [number of subjects and gestures] [Paper from Amma]
- CapgMyo - 23 participants and [number of gestures] gestures to classify (Du et al. 2017)

REFERENCES

1. How to Classify Works Using ACMs Computing Classification System. http://www.acm.org/class/how_to_use.html.

2. Badenov, B. Effects of prolonged use of WIMP user interfaces on Alces americana and Glaucomys volans. In *Proceedings of UIST '87* (February 30–April 1, Graceland, TN), ACM, NY, 1987, pp. 231–240.

3. Henry, T.R., Yeatts, A.K., Hudson, S.E., Myers, B.A., and Feiner, S.K. A nose gesture interface device: Extending virtual realities. *Presence 1*, 2 (Spring 1992), 258–261.

4. Schwartz, M. Guidelines for Bias-Free Writing. Indiana University Press, Bloomington, IN, USA, 1995.

5. Zaranka, W., Ed. *The Brand-X Anthology of Poetry: A Parody Anthology.* Apple-wood Books, Cambridge, MA, 1981.