

MIT Open Access Articles

*EMG-based Real Time Facial Gesture
Recognition for Stress Monitoring*

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

Citation: S. Orguc, H. S. Khurana, K. M. Stankovic, H. S. Leel and A. P. Chandrakasan, "EMG-based Real Time Facial Gesture Recognition for Stress Monitoring," 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Honolulu, HI, 2018, pp. 2651-2654

As Published: <http://dx.doi.org/10.1109/embc.2018.8512781>

Publisher: IEEE

Persistent URL: <https://hdl.handle.net/1721.1/123872>

Version: Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

Terms of use: Creative Commons Attribution-Noncommercial-Share Alike



EMG-based Real Time Facial Gesture Recognition for Stress Monitoring

S. Orguc¹, H. S. Khurana¹, K. M. Stankovic^{2,3}, H.S. Lee¹ and A.P. Chandrakasan¹

Abstract— An electromyogram (EMG) signal acquisition system capable of real time classification of several facial gestures is presented. The training data consist of the facial EMG collected from 10 individuals (5 female/5 male). A custom-designed sensor interface integrated circuit (IC) consisting of an amplifier and an ADC, implemented in 65nm CMOS technology, has been used for signal acquisition [1]. It consumes 3.8nW power from a 0.3V battery. Feature extraction and classification is performed in software every 300ms to give real-time feedback to the user. Discrete wavelet transforms (DWT) are used for feature extraction in the time-frequency domain. The dimensionality of the feature vector is reduced by selecting specific wavelet decomposition levels without compromising the accuracy, which reduces the computation cost of feature extraction in embedded implementations. A support vector machine (SVM) is used for the classification. Overall, the system is capable of identifying several jaw movements such as clenching, opening the jaw and resting in real-time from a single channel EMG data, which makes the system suitable for providing biofeedback during sleeping and awake states for stress monitoring, bruxism, and several orthodontic applications such as temporomandibular joint disorder (TMJD).

I. INTRODUCTION

In the recent years, there has been a growing interest in designing personalized devices for monitoring the physical and emotional wellbeing of individuals. Wearable and cost-effective medical devices that are able to extract information from certain physiological biomarkers in a non-invasive way to provide biofeedback and early diagnostic information can contribute to the development of the future of personalized medicine.

Stress monitoring is one of the emerging areas, where researchers have been exploring different biomarkers to measure the stress level of individuals. Examples include detecting the electrodermal activity (EDA) from skin to measure the change in resistance with sweating, or detecting the heart rate variability (HRV) in response to increased stress levels [2,3]. Electromyography (EMG), which is a tool to record and evaluate the electrical activity produced by our skeletal muscles, is a potentially useful technique for this application. EMG has been extensively used in

polysomnography (PSG) studies to evaluate the quality of sleep [4,5] and can provide information about the level of stress by recording from facial expressions during sleep to detect certain patterns in specific movements such as clenching. In addition to sleep studies, personalized EMG sensors capable of real-time decision making can be used to take measurements during daily activities in order to warn the user when an alarm event occurs. In the future, we envision wearable EMG devices that also have stimulation capability in order to provide relaxation to the corresponding muscle of interest.

However, there are two main challenges in this roadmap. First, EMG signals have extremely weak time-domain characteristics and the system requirements for signal acquisition is challenging. The signals must be captured from noisy electrode-skin interface with high signal-to-noise-ratio (SNR) while rejecting the common-mode interference as well as consuming low power for practical use. Hence, customized sensor interfaces that are compact and low power are required. Second, in order to classify certain patterns from the measurement, the signal behavior must be well understood and the systems must be trained with carefully labeled data.

Most of the previous work in real time EMG classification has been on hand and arm gestures, in order to control prosthetic devices [6-8]. There have been few studies on facial gesture recognition to identify several facial movements using commercial devices and electrodes at multiple locations on face, and neural networks and hyper dimensional computing (HDC) are used as feature extraction methods [9,10]. Customized devices with machine learning processor have been proposed in ECG-EEG domain with on-chip machine learning classifier for detection of seizures and cardiac-arrhythmia applications [11,12]. There is a potential to explore EMG sensor interfaces with embedded machine learning processors that can be used for several biofeedback applications such as stress monitoring.

This work presents an EMG-based facial gesture recognition system that can classify different jaw movements such as clenching, jaw opening, chewing and resting in real-time. The system is trained on data collected from 10 subjects in different recording sessions. The custom-designed analog front end (AFE) IC consists of an amplifier and an 8-bit SAR ADC, with a total power consumption of 3.8nW from a 0.3V power supply. The algorithm implements modifications on the 4-stage DWT decomposition to improve computational cost in feature extraction. It performs feature extraction and classification every 300ms for a real-time constraint of an engineering application [13].

*Research supported by Delta Electronics, Siebel Fellowship, MIT Center for Integrated Circuits and Systems, Lauer Tinnitus Research Center and TSMC University Shuttle Program.

S. Orguc¹, H.S. Khurana¹, H.S. Lee¹ and A. P. Chandrakasan¹ are with the Microsystems Technology Laboratories, Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: sirma@mit.edu).

K. M. Stankovic^{2,3} is with Harvard Medical School, Boston, MA 02115 USA, and also with Massachusetts Eye and Ear Infirmary, Boston, MA 02114 USA.

II. METHODS

A. System Setup

Fig. 1 shows the system setup. Two surface rounded pregelled Ag/AgCl electrodes that are 2 cm apart are placed on the opposite ends of the left masseter muscle to form a bipolar configuration, while the third one is placed to the forehead as a reference. During the recordings, the signal acquisition IC captures the single channel surface EMG signals from the electrodes and the FPGA transfers the digitized data to PC for data processing. We perform off-line training through feature extraction and classification in MATLAB.

The experimental procedures involving human subjects described in this paper were approved by the Institutional Review Board. The dataset is based on a collection of facial EMG recordings from 10 subjects (5 male / 5 female, age-range: 25-31). For each of the selected gestures, there is a designated button on the FPGA for the user to press while performing a particular activity in order to manually label the training data. A total of 20 recordings per 10 s are collected from each user.

B. Data Acquisition

The AFE is specifically designed to capture surface EMG signals. It consists of an amplifier and an 8-bit, 1kS/s SAR-ADC, which operate from 0.3V supply as shown in Fig. 2.a) [1]. The unique amplifier topology presents fewer cascaded transistors at each stage of amplification, hence allows reliable operation from low voltages. It provides a noise level of $26\mu\text{V}_{\text{rms}}$, 40dB gain and a state-of-the-art power efficiency factor (PEF) of 2.2 from a 20-425Hz bandwidth. High-pass corner frequency of 20Hz, which is optimal for the EMG signal content, substantially removes

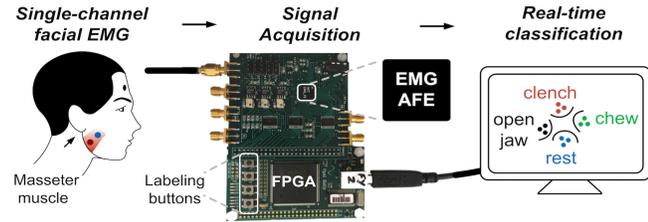


Figure 1. System setup.

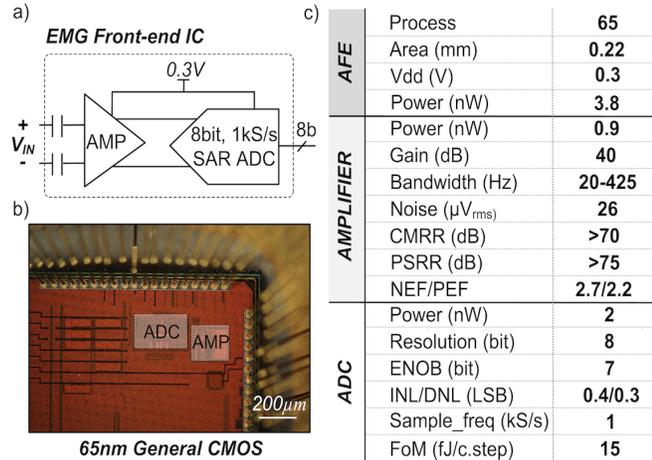


Figure 2. AFE details. a) Block diagram. b) Die photo. c) Performance.

the motion artifacts as well as the $1/f$ noise [14]. Together with the ADC, the prototype implemented in 65nm CMOS process consumes 3.8nW and has an area of 0.22mm^2 . Fig. 2 b) and c) show the die photo together with the performance summary of the sensor interface. The details of the AFE can be found in [1].

C. Signal Preconditioning

Although the data for each individual subject was collected in a single session and from the same muscle, the differences across the subjects (such as due to the facial anatomy, electrode displacement and muscle strength) results in different EMG characteristics for each user, which requires normalization prior to feature extraction. The normalization can be achieved by dividing the EMG activity during a task by a reference obtained from the same muscle, where the reference should have high repeatability [15]. In our study, the EMG data is normalized to the mean value during an activity for each activity and each individual separately.

D. Feature Extraction and Classification

Fast Fourier Transform (FFT) has been widely used in signal processing studies in order to extract the frequency domain information. However, it does not provide any time domain information, thus is not ideal for activity localization that is required for our application. In contrast, wavelet transform is found to be more suitable for non-stationary signals such as EMG, since it is a time-frequency based representation. The major hurdle in using wavelet transform is the high-dimensionality of the feature vectors, which can be computationally intensive.

DWT, which is the computationally practical form of the continuous wavelet transform (CWT), enables variable resolution in time-frequency domain. Fig. 3.a) shows the DWT structure used in this study, which uses 7th-order Daubechies wavelets (db_7) with 4 levels of decomposition that has been found suitable for EMG classification [7]. It first divides the signal into low-pass (LP) and high-pass (HP) branches by using the corresponding decomposition filters. The LP branch goes deeper as the signal gets down-sampled and divided into LP and HP branches at each stage to get the higher order coefficients. As a result, low frequencies have better frequency resolution, whereas the high frequencies have better resolution in time, represented by the coefficient subsets (CD_1 - CD_5).

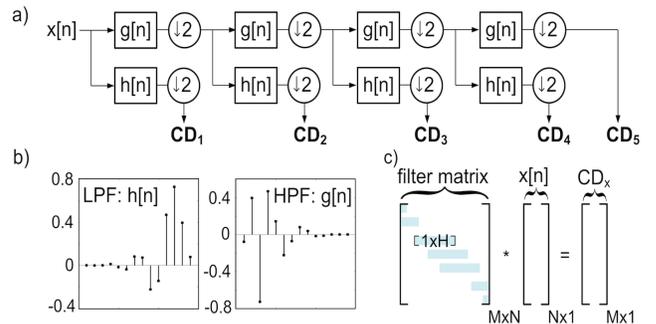


Figure 3. DWT transform. a) The decomposition tree for 4 stages. b) The Daubechies LPF and HPF coefficients. c) Visualization of the convolution matrices to get the DWT coefficients CD_x .

The required number of filtering and down sampling in order to get the subset coefficients is different for each branch, which results in different computational cost. For a given initial filter length of h , the equivalent filter length H for each branch can be found using the following formula:

$$H = \begin{cases} h + (h - 1) \sum_{i=1}^{N-1} 2^i & \text{if } b > 1 \\ h & \text{if } b = 1 \end{cases} \quad (1)$$

where, b is the number of the branch and N represents the number of un-sampling required for the corresponding branch. Even though the deeper branches have longer filter length, convolution matrix that consists of the shifted versions of elements of the equivalent filter has fewer components due to the increased number of un-sampling. The required number of rows (M) in the filter matrix for the convolution can be found as:

$$M = \frac{(H + X)}{2^{b-1}} \quad (2)$$

where, X is the input signal length. Hence, it is computationally less expensive to compute the higher order coefficients. Table I shows the values of H and M as well as the total number of multiply-accumulate operation (MAC) required to compute the subset coefficients for the chosen filter length of 14 and for an input signal of 300 samples. As seen from the table, CD_4 and CD_5 are potential candidates to use for the simplified DWT feature extraction, which we use in our final classification algorithm.

Fig. 4 shows the flow graph for feature extraction and classification. Each classification loop takes a 300-sample signal segment as the input (300ms window at 1kHz sampling rate). The feature extraction step first generates the selected DWT coefficients CD_4 and CD_5 as well as the moving average value (MAV) of the time domain signal (MAV_x). Then it computes the MAV and standard deviation (STD) of the subset coefficients for the selected branches using the following formula:

$$MAV_{CD_i} = \frac{1}{M} \sum_{k=1}^M |CD_i(k)| \quad (3)$$

where M is the total number of coefficients at each DWT branch, which was previously calculated in Table I. The STD of the DWT coefficients can be calculated as follows:

$$STD_{CD_i} = \sqrt{\frac{1}{M-1} \sum_{k=1}^M |CD_i(k) - MAV_{CD_i}|^2} \quad (4)$$

Hence, each window is represented by a feature vector that consists of 5 elements: MAV_{CD_4} , STD_{CD_4} , MAV_{CD_5} , STD_{CD_5} and MAV_x .

The algorithm uses support vector machine (SVM) for multi-class classification of the facial gestures. It first partitions the data into training and test sets. During model training, the algorithm implements 10% cross-validation on the training set. Once the system is trained offline, the accuracy is tested on the test set. For real-time classification, the model parameters of the trained classifier together with the pre-computed DWT filter matrices are used.

TABLE I. THE VALUES OF THE REQUIRED FEATURE EXTRACTION PARAMETERS FOR A GIVEN FILTER AND SIGNAL LENGTH

	H	M	# of MAC
CD_1	14	157	47100
CD_2	40	85	25500
CD_3	92	49	14700
CD_4	196	31	9300
CD_5	196	31	9300

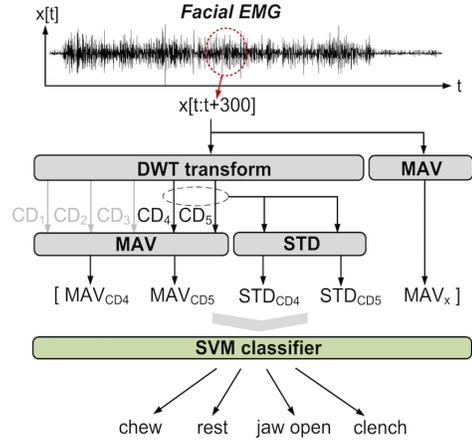


Figure 4. Flow chart summarizing the feature extraction and classification.

III. RESULTS AND DISCUSSIONS

Fig. 5 shows sample signal segments for the activities of interest together with the mean and standard deviation of the signal window (x), CD_4 and CD_5 for all subjects in order to visualize the potentials of the selected features in classification. The time domain averages of the activities can be used as features but are not sufficient for classification since the subjects have quite different muscle strength as it is reflected to the standard deviation.

Fig. 6 shows summarizes the system accuracy, where the algorithm uses one subject's data as the test data at each repetition of the training process to evaluate system performance on unseen user data. As seen, it can classify resting, clenching and chewing activities with $> 85\%$ accuracy. The reason for the lower accuracy in jaw opening classification results from similar feature values for rest and jaw opening as shown in Fig 5.b).

Fig. 7 compares our algorithm with the FFT computation as well as the 4-stage DWT algorithm, which uses all stage coefficients as its features. The feature vector for the FFT is constructed by calculating the $N=512$ point FFT for a 300ms window, which costs $N/2 \cdot \log_2 N$ complex multiplications and $N/2 \cdot \log_2 N$ complex additions. The average energy values for 10 bins are used to obtain the elements of the feature vector. As can be seen from the comparison, although it is computationally more efficient to use FFT, the DWT is superior in terms of the accuracy. On the other hand, in 4-stage decomposition, using all the DWT coefficients does not contribute to the accuracy significantly.

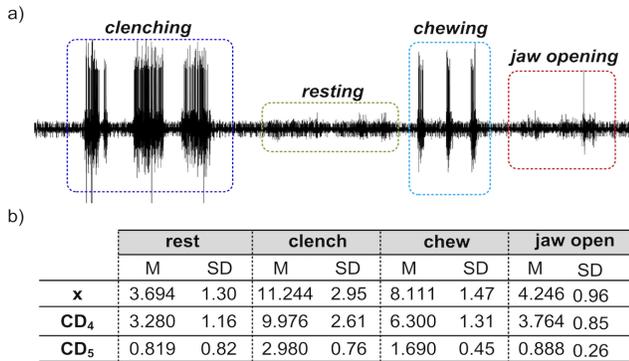


Figure 5. a) Representative signal segments. b) The mean (M) and standard deviation (SD) of the signal window as well as the selected DWT coefficients.

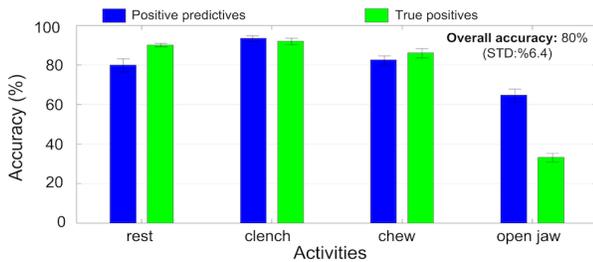


Figure 6. The true positive and positive predictive values for the activity detection. The algorithm is trained and tested 10 times, where at each repetition, the data from one subject is used as the test data.

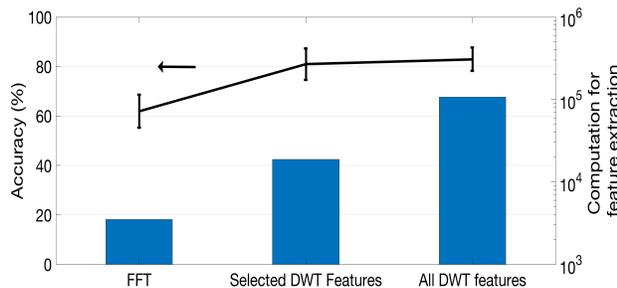


Figure 7. The number of computations required for feature extraction and the classification accuracy for 3 transforms: FFT, the selected DWT branches that our algorithm use, and the standard 4-stage DWT.

IV. CONCLUSION

We present real-time classification from single channel EMG for several facial gestures using DWT for feature extraction. We use an ultra low power sensor interface specifically designed for this application. Comparing the cost versus performance, the FFT operations require less computation per feature extraction, however the classification results are not as accurate as the DWTs. While DWT feature extraction is computationally more intensive, we reduce the computational complexity substantially by selecting only two branches, which require least number of computations, yet represent the signal as well as the full feature vector without compromising the classification accuracy. We believe that, although computationally more intensive than FFTs, the DWTs can be used in embedded EMG processing devices by improving the feature extraction process. Our system proposes a way to reduce the number of

computations to a level that is modestly higher than the FFT computations, while the performance is superior to FFTs due to the non-stationary behavior of the EMG characteristics. In the future, we envision that the wavelet transforms can be used in embedded EMG processors to enable low-cost real-time decision making for diagnostic as well as therapeutic purposes.

ACKNOWLEDGMENT

The authors acknowledge the funding support of Delta Electronics, Siebel Fellowship, MIT Center for Integrated Circuits and Systems, Lauer Tinnitus Research Center and the TSMC University Shuttle Program for chip fabrication.

REFERENCES

- [1] S. Orguc, H.S. Khurana, H.S. Lee, A.P. Chandrakasan, "0.3v ultra-low power sensor interface for emg", *European Solid State Circuits Conference (ESSCIRC) Dig. Tech. Papers*, Sep. 2017.
- [2] N. Attaran, J. Brooks, T. Mohsenin, "A low-power multi-physiological monitoring processor for stress detection", *IEEE Sensors Dig. Tech. Papers*, pp. 1-3, Oct. 2016.
- [3] C. Setz, B. Arnich, C. Schumm, R. La Marca, G. Troster, U. Ehlert, "Discriminating stress from cognitive load using a wearable eda device", *IEEE Trans. on Information Technology In Biomedicine*, vol. 14, no. 2, pp. 410-417, March 2010.
- [4] A. Chemparathy, H. Kassiri, M. T. Salam et. al., "Wearable low latency sleep stage classifier", *IEEE Biomedical Circuits and Systems Conference (BioCAS) Proceedings*, pp. 592-595, Oct. 2014.
- [5] P. Dreyer, W. Yachida, N. Huynh et. al., "How close can single-channel emg data come to psg scoring of rhythmic masticatory muscle activity?", *Journal of Dental Sleep Medicine*, vol. 2, no. 4, pp. 147-156, 2015.
- [6] A. Rahimi, S. Benatti, P. Kanerva et. al., "Hyperdimensional biosignal processing: a case study for emg-based hand gesture recognition", *IEEE International Conference on Rebooting Computing (ICRC)*, pp. 1-8, 2016.
- [7] A. Phinyomark, C. Limsakul, P. Phukpattaranont, "Application of wavelet analysis in emg feature extraction for pattern classification", *Measurement Science Review*, vol. 11, no. 2, pp. 45-52, 2011.
- [8] S. Benatti, F. Casamassima, B. Milosevic et. al., "A versatile embedded platform for emg acquisition and gesture recognition", *IEEE Transactions On Biomedical Circuits And Systems*, vol. 9, no. 5, pp. 620-630, 2015.
- [9] A. Gruebler, K. Suzuki, "Design of a wearable device for reading positive expressions from facial emg signals", *IEEE Transactions On Affective Computing*, vol. 5, no. 3, pp. 227-237, 2014.
- [10] M. Hamed, S-H. Salleh, M. Astaraki and A. M. Noor, "Emg based facial gesture recognition through versatile elliptic basis function neural network," *Biomedical Engineering Online*, vol. 12, no. 1, pp. 73, 2013.
- [11] J. Zhang, L. Huang, Z. Wang, N. Verma, "A seizure detection IC employing machine learning to overcome data-conversion and analog-processing non-idealities", *IEEE Custom Integrated Circuits Conference (CICC)*, pp. 1-4, 2015.
- [12] K.H. Lee, N. Verma, "A low-power processor with configurable embedded machine-learning accelerators for high-order and adaptive analysis of medical-sensor signals", *IEEE Journal Of Solid-State Circuits*, vol. 48, no. 7, pp. 1625-1637, July 2013.
- [13] K. Englehart, B. Hudgins, P.A. Parker, "A wavelet-based continuous classification scheme for multifunction myoelectric control", *IEEE Trans. Biomed. Eng.*, vol. 48, no. 3, pp. 302-311, 2001.
- [14] J.C de Luca, L.D. Gilmore, M. Kuznetsov, S. H. Roy, "Filtering the surface emg signal: movement artifact and baseline noise contamination," *J. of Biomechanics*, vol. 43, no.8, pp. 1573-1578, 2010.
- [15] M. Halaki, K. Ginn, "Normalization of emg signals: to normalize or not to normalize and what to normalize to?" in *Computational Intelligence in Electromyography Analysis - A Perspective on Current Applications and Future Challenges*. pp. 175-194.