2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)

Spontaneous and posed smile recognition based on spatial and temporal patterns of facial EMG

Monica Perusquía-Hernández
School of Integrative and Global Majors
University of Tsukuba
Email: monica@ai.iit.tsukuba.ac.jp

Mazakasu Hirokawa
Faculty of Engineering,
Information and Systems
University of Tsukuba
Email: hirokawa_m@ieee.org

Kenji Suzuki
Faculty of Engineering,
Information and Systems
University of Tsukuba
Email: kenji@ieee.org

***Abstract*—Facial expressions are one of the most salient cues of affect. However, they are also a communication tool that can be expressed voluntarily. Spontaneous and posed facial expressions have different characteristics. In this study, we show the potential of using Electromyography (EMG) in a wearable device to automatically differentiate between posed and spontaneous smiles. This classification is moderately successful when using spatial and magnitude features, and increases significantly if temporal features are included. Hence, the high temporal resolution of EMG-based detection is advantageous in this task.**

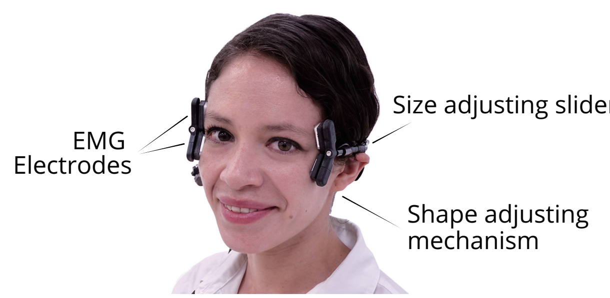


Figure 1: The wearable device used to record EMG during the elicited smiling behavior.

1. Introduction

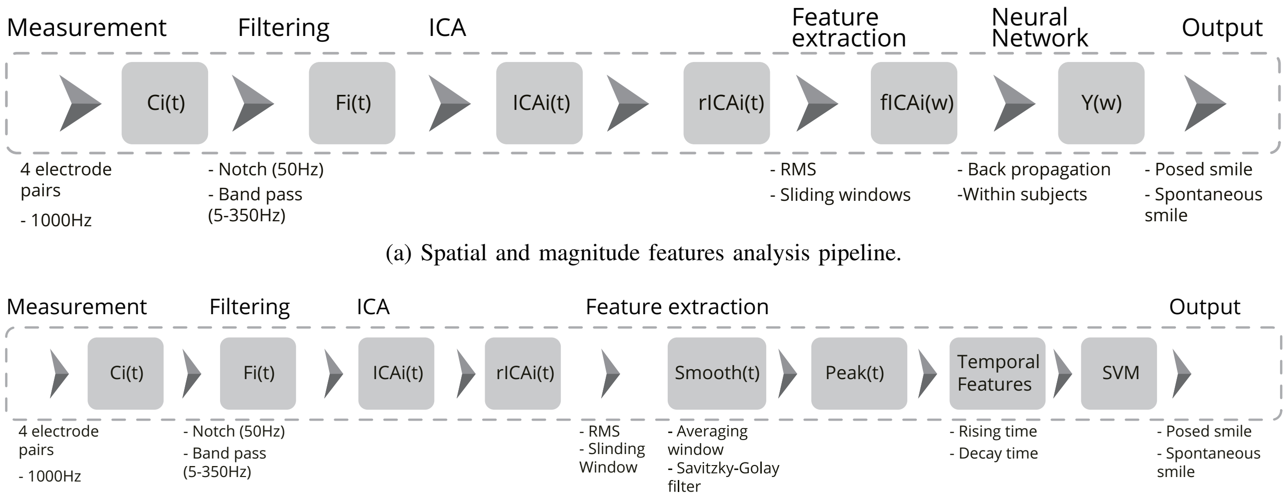
Automatic detection of human affect is beneficial in diverse applications. Knowing if a person is happy could help to assess its satisfaction with a service or a product; the quality of life of a patient who cannot communicate otherwise; or to create empathy in human-robot interaction.

One of the most used cues to detect affect is facial expression recognition. Facial expressions are usually related to emotional states of a person. However, these facial expressions can also be voluntarily fabricated to transmit a deliberate message. According to [1], "the movements inherent to posed facial expressions display an emotion an expresser ostensibly intends to convey, whereas spontaneous facial expressions correspond to an expressers actual, unmitigated emotional experiences." Moreover, spontaneous facial expressions are an automatic motor movement, and posed facial expressions are voluntary. Hence, they are believed to have different neural pathways [2], and they are represented by different temporal dynamics from Electroencephalography (EEG) signals [3], [4].

Perhaps the most commonly studied facial expression regarding posed and spontaneous differences are smiles. Besides expressing happiness, a smile can also be used to convey kindness to others. Different terms have been used to refer to these smile-types. Posed and deliberate smiles are often used as synonymous, and they are opposite to spontaneous smiles. Several differences among posed and spontaneous smiles have been found. The most sound difference is the activation of the *orbicularis oculi* muscle that was believed to happen during spontaneous smiles only [5]; in the so-called Duchenne smile. Nevertheless, recent studies have found that this muscle is activated both in posed and spontaneous smiles [1]. Furthermore, posed smiles tend to have a larger amplitude [2], [6], [7]. Besides these spatial differences, spontaneous and posed facial expressions differ substantially in their temporal dynamics [2], [6], [7], [8], [9], [10]. Whilst these vary, most agree that posed and spontaneous smiles differ in amplitude, rising and decaying speed, and duration. Spontaneous smiles tend to last longer than posed ones [7], [9], [10]; they have multiple peaks [10]; and they have longer rising, decaying, and peak durations [6], [10]. Furthermore, posed smiles have a longer onset and offset speed [6]. According to [5], spontaneous expressions have a fast and smooth onset; with apex coordination, in which muscle contractions in different parts of the face peak at the same time. In posed expressions, the onset tends to be slow and jerky, and the muscle contractions typically do not peak simultaneously.

Most of these evidence has been found using human coding and Computer Vision methods [1], [2], [10]. However, these have limitations in the time needed to process the data before differentiating these smiles. Furthermore, their sampling rate is limited. In contrast, physiological signals have a higher temporal resolution and are able to pick up information even if it is not visually perceivable. Several studies have already proven the feasibility of recognizing

978-1-5386-0563-9/17/$31.00



(a) Spatial and magnitude features analysis pipeline.

(b) Temporal feature analysis pipeline.

Figure 2: The signal processing steps for each type of analysis.

emotion in a real-time setting from electro-physiological signals. For example, [11] used Electromyography (EMG), and skin conductance to add real-time affection detection to a gaming scenario. Other studies have also explored emotion recognition using EMG. [12] performed emotion recognition from EMG using a wavelet transform and different pattern recognition methods, where Support Vector Machines (SVM) outperformed Back-Propagation (BP) and Lagrange-Multiplier (L-M) improved BP Neural Networks. However, these were used to detect emotion categories and not specific facial expressions.

In this study, we propose to use EMG wearable technology to explore the temporal characteristics of spontaneous and posed similes. In the following sections, the methodology to collect samples of these smiles is described, along with the wearable used to measure distal EMG. Finally, the data analysis to unveil differences among these two types of expression is included.

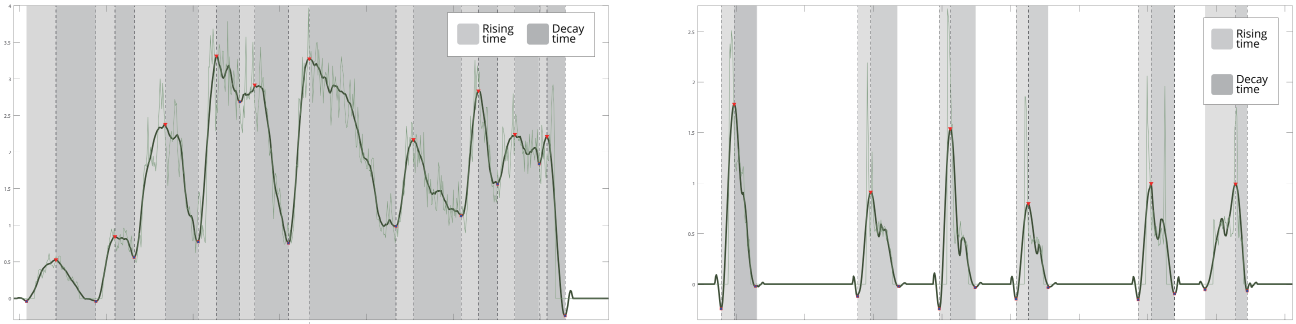
2. Wearable EMG

Four wireless surface dry-electrode EMG channels from Biolog were used to record the data. The electrodes were placed on the side of the face, on top of the *temporalis* and the *zygomaticus major* muscles. This arrangement was chosen following the guidelines from previous research [13]. We use the distal electrode locations on the side of the face in order to capture facial expressions. Hence, we do not identify the activity of each facial muscle, but we conduct a pattern classification where facial expressions are considered as a combination of the activity of all related facial muscles. Due to the overlapping of the distal signal from different muscle groups, special biosignal processing is applied.

This configuration was proven to be robust for detection of smiles and frowns in various settings [14], [15], [16]. These four electrodes were placed in a circlet that can be comfortably worn without covering the face (Figure 1).

3. Data collection

1) **Experiment design and procedure.** Participants were informed that the purpose of the test was to rate some videos with a questionnaire and by measuring their facial EMG. To elicit spontaneous smiles, a series of fun video stimuli were showed to the participants in a counterbalanced order. In total, participants watched eight videos. During the first four, they were asked to "keep a neutral face while watching the videos". Therefore, we expected all leaked expressions to be spontaneous. In the second block, no particular instruction was provided. Before and after watching the stimuli, participants were asked to pose smiles, frowns, and eye-brow lifts, supposedly with the purpose of verifying the EMG signal. After the experiment, participants were debriefed.
2) **Stimuli.** Three Ad videos known to elicit smiles were selected from previous research [17]: "The force" (Video TF, 62s), "House sitting" (Video HS, 30s), "Parisian Love" (Video PL, 53s). Additionally, an edition of the 2011 Jimmy Kimmel Challenge "I Told My Kids I Ate All Their Halloween Candy" (Video HA, 2min 9s), was included. During these four videos, participants were asked to avoid making any facial expressions. Four extra videos showing fun and cute behavior were watched with no particular instruction: "Baby expectancy" (Video 5, 29s), "Fun kiddies" (Video 6, 2min 18s), "Brotherhood" (Video 7, 53s), "Dirt Devil" (Video 8, 1min 28s). All videos were presented at 30 frames per second with 720x480 pixel resolution.
3) **Participants.** Sixteen voluntary participants took part on the study (average age=26.3 years old, SD=3.24, 6 female). Eight participants were Japanese and the rest from other European and



(a) Spontaneous smiles. The X-axis represents sample number at 100 Hz. The Y-axis is the smoothed absolute value of one of the ICA components of participant 16.

(b) Posed smiles. The X-axis represents sample number at 100 Hz. The Y-axis is the smoothed absolute value of one of the ICA components of participant 16.

Figure 3: A sample of smile data with the estimated envelopes, peaks, and rise and decay sections. No-smile EMG data was masked with zeros for easy visualization.

Latin American countries. None of them had experience using the measuring device, and 11 participants had seen at least one of the videos before.

4) **Measurements.** During the task, surface EMG and the participant's face were recorded simultaneously. The surface EMG was recorded at 1 kHz sampling rate using a four channel Biolog DL-4000 system. The camera used was a Sony Cyber-shot DSC-RX100 II with 1920 x 1020 resolution at 120 fps.

4. Data analysis

4.1. Video coding

All the recordings of the participants' face while watching the stimuli were coded frame-by-frame for facial expressions by two experienced coders. The labeling included coding for the onset, offset, and apex frames of the facial expression; the Facial Action Unit Systems (FACS) Action Units that were present in the expression; and whether it was considered a smile or not, a posed expression or not, and laughter or not. Smiles were often a display of AU6, AU12 and/or AU25. However, the smile label was not assigned every time these AU occurred [17]. All facial movements considered as swallowing, coughing, or sneezing were excluded. For the posed expressions block, the instruction given to the participant was used to label the expression as a smile or not. Furthermore, an experienced coder labeled the data in the same manner as described for the stimuli block to identify the start and the end of the posed facial expressions.

4.2. Signal processing

The EMG signals picked up by the electrodes are transmitted to a laptop via Bluetooth, where they are analyzed using Matlab 2014a. Two different detection algorithms are proposed. The first one relies on magnitude and spatial features only, whereas in the second, temporal features of the signal were included.

4.2.1. Spatial and magnitude features analysis pipeline. The data from all four channels is band-pass filtered from 5 to 350 Hz. Second, it is notch filtered at 50 Hz and its harmonics up to 350 Hz. Afterwards, any linear trends were removed from the signal. Next, Independent Component Analysis (ICA) was applied, to separate the components from different muscles. Then, the absolute value of the resulting components was calculated. Next, the RMS value is calculated over overlapping windows of 100 ms, sliding one sample at a time. The resulting data was labeled according to the human coding, and used as features to train a Neural Network (NN) with one hidden layer of four Sigmoid neurons. To validate this model, cross-validation with 70% train, 15% validation, and 15% test data was used. The neural network aims to compare spontaneous smiles with posed smiles. Due to the unbalanced nature of the data, the majority class was undersampled to match the samples of the minority class [18], [19]. This process is shown in figure 2a.

4.2.2. Temporal feature analysis pipeline. This method followed a similar preprocessing as the previous one (Figure 2b). The four channels were band-pass filtered (5-350 Hz); notch filtered (50 Hz and harmonics); de-trended; the absolute value of its ICAs was calculated; and the RMS of the signal was calculated using an overlapping 100 ms window sliding every sample. Afterwards, smile data was selected according to the human label, and sliced in individual smiles. Next, each smile data was smoothed using an averaging non-overlapping window of 100 ms, and a Savitzky-Golay Filter with a 5th order polynomial and 41 as frame length. Then, peak detection was performed on the smoothed EMG signal to calculate the rising, and decay times (Figure 3). The rising time is defined as the time taken from the first minimum to the first maximum (peak) in the smoothed signal; decay time is defined as the time between the last maximum to the last minimum. Furthermore, the magnitude change during rising and decay; and the rising and decay speed were calculated as features. The resulting feature set was standardized and used to train a Support

TABLE 1: Spontaneous and posed smile detection results.

(a) EMG-based posed and spontaneous smiles identification. RMS was used as feature as described in section 4.2.1

Participant	Spontaneous-smile detection		
	Precision	Recall	Accuracy
1	75.20%	85.90%	81.40%
2	68.00%	64.90%	65.60%
3	78.90%	77.20%	77.80%
4	68.50%	80.00%	75.70%
5	91.80%	93.30%	92.60%
6	83.00%	72.40%	75.70%
7	78.60%	70.90%	73.10%
8	98.50%	97.20%	97.80%
9	33.00%	61.30%	56.10%
10	80.50%	71.20%	73.90%
11	71.00%	74.40%	73.30%
12	94.90%	88.50%	91.30%
13	92.00%	80.50%	84.80%
14	92.20%	87.50%	89.50%
15	83.70%	76.70%	79.10%
16	73.80%	81.60%	78.60%

(b) EMG-based posed and spontaneous smiles identification. Spatio-temporal features were used as described in section 4.2.2

Participant	Spontaneous-smile detection		
	Precision	Recall	Accuracy
1	85.16%	97.32%	87.98%
2	90.00%	100.00%	90.70%
3	97.65%	94.32%	95.33%
4	87.50%	78.87%	85.63%
5	96.30%	100.00%	96.43%
6	90.44%	86.62%	88.45%
7	87.94%	100.00%	89.43%
8	92.05%	99.29%	92.74%
9	91.84%	60.81%	85.23%
10	91.67%	72.13%	88.40%
11	86.13%	93.13%	88.18%
12	90.50%	87.10%	88.67%
13	84.75%	89.82%	86.54%
14	87.55%	90.27%	88.08%
15	89.37%	93.91%	89.41%
16	84.62%	84.62%	86.29%

Vector Machine (SVM) to distinguish between posed and spontaneous smiles. A Gaussian Kernel Function was used. To validate the model, a cross-validation with 70% train, 15% validation, and 15% test data was used. As with the other method, the data was balanced to match the minority class.

5. Results

In total, 240 spontaneous smiles and 353 posed smiles were identified by the human coders. The Cohens Kappa Coefficient of inter-rater agreement for spontaneous smile labeling was 0.41 ($p < 0.01$). The balance between the two types of facial expression depended heavily on the manner in which individual participants responded to the stimuli. However, we could get at least nine spontaneous smiles from each participant, and at most 19.

The spatial and magnitude features analysis pipeline was moderately successful in distinguishing spontaneous and posed smiles. Table 1a shows the classification results. These range from 56.10% till 97.80% of accuracy. On the other hand, the temporal feature analysis pipeline (table 1b) accuracy values ranged from 85.23% till 96.43%. This classification was made in an average of 350 temporal features per participant extracted from the envelopes of the EMG measured while smiling. From these features, a series of t-tests revealed that spontaneous smile duration differs from posed smile duration ($t(2276)=-11.535$, $p<0.01$). Second, the magnitude both types of smiles is not significantly different ($t(2276)=-0.19837$, $p>0.05$). Finally, we observed that the rising time ($t(1151)=-7.5336$, $p<0.01$) and decay time ($t(1124)=-8.8359$, $p<0.01$) differ, but the speed of change is not significant. Neither during the rising phase ($t(1130)=0.22068$, $p>0.05$) nor during the decaying phase ($t(1108)=1.6413$, $p>0.05$).

6. Discussion and conclusions

EMG has the potential to distinguish between spontaneous and posed smiles in a portable and real-time fashion using a wearable. In this study, we provided support for this using both magnitude and temporal features. Our proposed algorithms take advantage of the ICA extraction to estimate different sources of the EMG signal and its magnitude. Furthermore, they also profit from the EMG's high temporal resolution to estimate smile characteristics without consuming excessive computational resources. From these two alternatives, the most successful results were given by considering the temporal resolution of the signal. As supported by previous studies, this is probably because spontaneous smiles and posed smiles differ in this aspect. In our data, the main difference was that spontaneous smiles tend to last longer than posed ones. Nevertheless, this might be influenced by the duration of the instruction to pose a smile, and the method used to elicit spontaneous expressions. Therefore, the nature of the smiles needs to be interpreted carefully according to each particular context and application. Posed smiles might last longer when they are intended to convey a positive message to an interlocutor; and spontaneous smiles might be as short as a quarter of a second, as in the case of micro-expressions [20]. In future work, we would like to explore in more detail how these temporal characteristics differ by the communicative intention of the wearer, and due to cultural differences.

References

[1] S. Namba, S. Makihara, R. S. Kabir, M. Miyatani, and T. Nakao, “Spontaneous Facial Expressions Are Different from Posed Facial Expressions: Morphological Properties and Dynamic Sequences,” pp. 1–13, 2016.

[2] J. F. Cohn and K. Schmidt, “The timing of facial motion in posed and spontaneous smiles,” *International Journal of Wavelets, Multiresolution and Information Processing*, vol. 2, pp. 121–132, 2004.

[3] G. Recio, O. Shmuilovich, and W. Sommer, “Should I smile or should I frown? An ERP study on the voluntary control of emotion-related facial expressions,” *Psychophysiology*, vol. 51, no. 8, pp. 789–799, 2014.

[4] A. Gentsch, C. Weiss, S. Spengler, M. Synofzik, and S. Schütz-Bosbach, “Doing good or bad: How interactions between action and emotion expectations shape the sense of agency,” *Social Neuroscience*, vol. 10, no. 4, pp. 1–13, 2015.

[5] P. Ekman and E. Rosenberg, *What the face reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS)*, second edi ed., P. Ekman and E. Rosenberg, Eds. Oxford University Press, 2005.

[6] K. L. Schmidt, Z. Ambadar, J. F. Cohn, and L. I. Reed, “Movement differences between deliberate and spontaneous facial expressions: Zygomaticus major action in smiling,” *Journal of Nonverbal Behavior*, vol. 30, no. 1, pp. 37–52, 2006.

[7] K. Schmidt, S. Bhattacharya, and R. Denlinger, “Comparison of deliberate and spontaneous facial movement in smiles and eyebrow raises,” *Nonverbal Behaviour*, vol. 33, no. 1, pp. 35–45, 2009.

[8] M. F. Valstar, H. Gunes, and M. Pantic, “How to distinguish posed from spontaneous smiles using geometric features,” *International Conf. Multimodal Interfaces*, vol. terfaces, p. 3845, 2007.

[9] M. Hoque, L. P. Morency, and R. W. Picard, “Are you friendly or just polite? - Analysis of smiles in spontaneous face-to-face interactions,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 6974 LNCS, no. PART 1, 2011, pp. 135–144.

[10] M. Mavadati, P. Sanger, M. H. Mahoor, and S. Y. Street, “Extended DISFA Dataset: Investigating Posed and Spontaneous Facial Expressions,” pp. 1–8, 2016.

[11] A. Nakasone, H. Prendinger, and M. Ishizuka, “Emotion recognition from electromyography and skin conductance,” in *Proceedings 5th International Workshop on Biosignal Interpretation*, no. January, 2005, pp. 219–222.

[12] S. Yang and G. Yang, “Emotion Recognition of EMG Based on Improved L-M BP Neural Network and SVM,” vol. 6, no. 8, pp. 1529–1536, 2011.

[13] A. Gruebler and K. Suzuki, “Design of a Wearable Device for Reading Positive Expressions from Facial EMG Signals,” *IEEE Transactions on Affective Computing*, vol. PP, no. 99, pp. 1–1, 2014.

[14] M. Hirokawa, A. Funahashi, Y. Itoh, and K. Suzuki, “A Doll-type Interface for Real-time Humanoid Teleoperation in Robot-Assisted Activity,” in *IEEE International Symposium on Robot and Human Interactive Communication*, Edinburgh, 2014, pp. 174–175.

[15] Y. Takano and K. Suzuki, “Affective communication aid using wearable devices based on biosignals,” in *Proceedings of the 2014 conference on Interaction design and children - IDC ’14*. New York, New York, USA: ACM Press, 2014, pp. 213–216. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2593968.2610455

[16] A. Funahashi, A. Gruebler, T. Aoki, H. Kadone, and K. Suzuki, “Brief report: The smiles of a child with autism spectrum disorder during an animal-assisted activity may facilitate social positive behaviors - Quantitative analysis with smile-detecting interface,” *Journal of Autism and Developmental Disorders*, vol. 44, no. 3, pp. 685–693, 2014.

[17] D. McDuff, R. El Kaliouby, T. Senechal, M. Amr, J. F. Cohn, and R. Picard, “Affectiva-mit facial expression dataset (AM-FED): Naturalistic and spontaneous facial expressions collected ’in-the-wild’,” *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pp. 881–888, 2013.

[18] H. He and E. A. Garcia, “Learning from imbalanced data,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, no. 9, pp. 1263–1284, 2009.

[19] L. A. Jeni, J. F. Cohn, and F. De La Torre, “Facing Imbalanced Data Recommendations for the Use of Performance Metrics.” *International Conference on Affective Computing and Intelligent Interaction and workshops : [proceedings]. ACII (Conference)*, vol. 2013, pp. 245–251, 2013.

[20] P. Ekman and W. P. Friesen, “Nonverbal leakage and clues to deception,” *Psychiatry*, vol. 32, no. 1, pp. 88–106, 1969.