A Wearable Device for Fast and Subtle Spontaneous Smile Recognition

Monica Perusquía-Hernández, *Student Member, IEEE*, Mazakasu Hirokawa, *Member, IEEE*, and Kenji Suzuki, *Member, IEEE*

Abstract—Facial expressions are usually linked to emotional states of a person, and are among the most salient cues for automatic emotion recognition. They are an indispensable social communication tool, and therefore, they can also be fabricated to face complex situations in social interaction. Despite efforts to either deliberately or unconsciously conceal an emotion, micro-expressions are usually leaked. Because of their revealing nature, potential applications can arise from automatically detecting them, both at a personal and inter-personal levels. Therefore, we explored micro-expression detection using a wearable device that detects distal facial EMG signals unobtrusively, as an alternative to Computer Vision-based detection. EMG recognition is advantageous over commonly used video recognition techniques because of its robustness against occlusion and light changes, good temporal resolution, and independence of movement. We evaluated the performance of the wearable device on micro-smile recognition. The results show the potential of EMG to detect such fast and subtle spontaneous expressions. Finally, we evaluated the device as a tool to provide affective annotations of Advertisement Videos, and social interaction in a face-to-face context while watching these stimuli.

Index Terms—Electromyography, micro-expression recognition, wearable interface

1 Introduction

Emotional states and their expression take a central role in human life: they are indispensable for wellbeing, and for social communication and interactions. Those emotional states with positive affect are especially important. As argued by [1], "the pursuit of happiness is one of the most fundamental human motives". Most of our decisions and actions are driven by the affective forecast of what option will make us happier. Furthermore, expressing positive affect, via smiling and laughing, can strengthen the bond between two individuals [2]. Therefore, automatic recognition of behavioral cues of positive emotion is relevant for a number of application domains.

Recognition of these cues could support the fostering and assessment of the mental wellbeing and quality of life of a person or a group, especially of those with developmental disabilities that are unable to effectively communicate [3], [4]. They can also be used as motivational reward in training therapies and tutoring systems [5], [6]. They can be an input to create technology that adapts to human behavior and mental states, such as robots [7]; or to assess the effectiveness of products [8] and media [9], [10]. They can provide better understanding of users and patients during interviews, despite them giving socially desirable answers. This would help both designers of technology and caregivers to adapt their choices to the needs of their users and patients, respectively. Also at a personal level, it would be useful to be aware of our own expressions. Such feedback could be used for social facial expression awareness training, to improve interpersonal communication. This might prove especially useful for persons with autism spectrum disorders (ASD), who have difficulties with social interactions, probably due to difficulties to recognize facial expressions of emotion [11], [12], [13]. Finally, it would also be helpful for blind people to perceive their own expressions [14] and the expressions of others. This would create bonding with their interlocutors, and support their social interactions.

Among the behavioral cues of positive affect, laughter is commonly targeted as the communicative signal of enjoyment per excellence. Spontaneous laughter is often described as if the people experiencing it abandon themselves to the bodily response of such enjoyment [15]. According to previous research, laughter is composed of respiration, vocalization, body movement, and facial action [15], [16]. The facial action is mainly that described as a Duchenne display, or a genuine smiling [15]. Laughter is then accompanied by a series of respiration, vocalization, and body movement bursts. These bursts are often referred as a laughter bout. While laughter is reported to have a mode of four pulses, laughter with one or two pulses also exists. One-bout laughter is called exclamation laughter or chuckle [15]. As the onset of laughter often presents a prevocal smiling expression [16], we argue that fast and subtle smiles are similar to a first laughter burst that is quickly contained. In this sense, automatically detecting such

- *M. Perusquía-Hernández is with the Department of Empowerment Informatics, University of Tsukuba, Tsukuba, Ibaraki 305-8577, Japan. E-mail: monica@ai.iit.tsukuba.ac.jp.*
- *M. Hirokawa and K. Suzuki are with the University of Tsukuba, Tsukuba, Ibaraki 305-8577, Japan. E-mail: hirokawa@emp.tsukuba.ac.jp, kenji@ieee.org.*

Manuscript received 30 July 2016; revised 31 Aug. 2017; accepted 10 Sept. 2017. Date of publication 25 Sept. 2017; date of current version 5 Dec. 2017. (Corresponding author: Monica Perusquía-Hernández.)
Recommended for acceptance by M. Mancini, R. Niewladomski, S. Hashimoto, M. E. Foster, S. Scherer, and G. Volpe.
For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.
Digital Object Identifier no. 10.1109/TAFFC.2017.2755040

facial expressions and laughter in a continuous manner requires the same basic principle.

Computer Vision (CV) techniques are commonly used to automatically recognize facial expressions. An alternative recognition method is Electromyography (EMG), but it has been less explored due to its lack of usability. Furthermore, most of the work already done has focused in posed facial expressions, which are different from spontaneous facial expressions. Recently, an effort has been done to bring these results outside of the laboratory. This implies a shift in research towards the identification of spontaneous expressions in more ecologically valid environments.

In the following sections, a brief review of facial expression recognition methods is provided, followed by the proposal and evaluation of an easy-to-use EMG-based wearable device to detect fast and subtle facial expressions of positive affect. Finally, we explored the case study of measuring human-human and human-machine interactions with the proposed method. The quantification of behavioral cues of smile patterns can be used to highlight frames in a certain video (i.e., stimuli), according to the facial expression behavior of people watching it. This could provide affective scene understanding. To the best of the authors' knowledge, this evaluation provides evidence on the potential of EMG to create practical spontaneous facial expression detection at micro-expression level for the first time. Because they last less than 0.5 s, micro-expressions are difficult to detect using conventional methods. Furthermore, micro-smile based laughter detection is a novel approach because the preferred modality for laughter detection is often sound or body movement [16].

2 Related Work

Facial expression is a basic method that humans have developed to communicate their emotions. They are a social tool that allows us to differentiate from pleasant situations and dangerous ones. For example, it is possible to decide when it is appropriate to make a joke based on whether our interlocutor is smiling or not. On the other hand, facial expressions can also be fabricated, to conceal the true emotions of a person. In certain situations, this is socially desirable, such as avoiding smiling during a serious talk. Therefore, they are among the most used behavioral cues for emotion recognition.

2.1 Computer Vision-Based Methods

In the last years, the scientific community has addressed the facial expression recognition challenge using mainly Computer Vision [17]. A number of methods have been developed to detect facial expressions in static images and videos. These methods include tracking of geometric features, often called facial feature points or landmarks; and appearance-based methods using features such as Gabor filters [18], [19], [20]. In a review of ten CV-based machine learning studies to recognize emotion, [21] found that most studies rely on facial expressions based on the Facial Action Coding System (FACS) [22]. These studies achieved between 72 and 98 percent of accuracy. However, the number of features extracted, emotion states detected, and number of subjects participating in the studies varied considerably.

The main advantages of camera-based detection are its high spatial resolution sensitivity; and the unobtrusiveness that can be achieved with the recordings. Despite these advantages, and although the achieved accuracy is relatively high, CV-based facial expression recognition in an uncontrolled environment is still a challenging task. Good recording conditions and having a canonical view of the face are often required [21], [23], [24]. This means that the face should be uncovered and that the person is not allowed to face down or to the sides, which is unnatural. Furthermore, these methods are not robust against occlusion and poor lighting conditions [25]. Recently, some researchers have suggested the use of near-infrared (NIR) video sequences for facial expression recognition to overcome lighting variance [26], [27]. However, occlusion, movement, and point of view challenges remain.

2.2 Electromyography-Based Methods

An alternative to camera-based recognition is Surface Electromyography-based recognition. EMG has been used extensively in basic research of different fields to quantify facial movement activation [12], [28], [29], [30]. Previously, several researchers attempted to use EMG for facial expression detection. In a review of ten EMG-based facial expression detection algorithms, [31] found out a classification performance ranging from 56 till 100 percent. These studies varied in the number of expressions detected, the number of EMG channels used, the features used, and the classifiers applied. Similar to CV-based methods, comparison among studies is difficult due to the different methodologies pursued. In their exploration, [31] reported a comparison of multiple time-domain EMG features, of which the best performance of 87.1 percent accuracy was achieved with the Maximum Peak Value of the EMG signal.

An advantage of EMG signal is that current EMG technology allows for wireless and compact detection. Therefore, surface EMG could be integrated in a wearable capable of detecting such facial expressions. In [32], an eyebrow emotional expression recognition using surface EMG was proposed. The authors designed a headband embedded with a 10-channel digital EMG. They used different Root Mean Square (RMS) features and an Elman Neural Network classifier to detect five posed facial expressions related to eyebrow movement. These included sadness, anger, surprise, disgust, and fear. The average accuracy was 96.12 percent. Another wearable designed to detect facial expressions with EMG was proposed by [33]. In this work, the authors evaluated different electrode positions to determine the most suitable locations of the electrode leads so that the wearable would meet certain requirements. The proposed requirements were (1) minimal volumetric and (2) surface displacement to avoid obtrusiveness and electrode movement; (3) high rectified EMG amplitude; and (4) high correlation between signals. Their results show an average precision of 98 percent for posed smile detection and 96 percent for posed frowns. Although they detected only two types of facial expressions, they managed to do so with high precision using distal EMG. In other words, the EMG electrodes do not have to be placed directly on top of the muscles activated during the facial expression. They can be placed on the sides of the face. In this manner, one of the critical disadvantages of EMG was overcome, namely, the inhibition of facial movements caused by extensive electrode application on the face [30]. Furthermore, this EMG-based detection technology has been exploited in a real-time online feedback setting [34]; and during Animal-assisted therapies for ASD children [35]. The latter is an example of a condition where high movement and occlusion happens. However, the online feedback provided has a limited temporal resolution constrained by a time window of 150 ms, and the nature of the vibration feedback used. Finally, they used prototypical and very visible posed smiles as training data set.

Despite the potential of EMG-based detection, another possible disadvantage is that the EMG magnitude might change over time as a function of fatigue [30]. On the other hand, facial EMG measurement techniques have the advantage of providing an instantaneous, fine-grained muscle activity detection. They are capable of detecting muscle contractions that are too fast or too small to be visually perceived [8], [30].

2.3 Spontaneous Facial Expression Detection

Most of the work previously done is about posed expressions, as opposed to spontaneous expressions. Posed facial expressions are facial expressions collected in a controlled environment when a subject is asked to deliberately produce them. Spontaneous facial expressions are those displayed by freely behaving individuals. These have different characteristics than posed expressions. They often involve different facial muscles; their temporal dynamics are different; and they are mediated by distinct neural pathways [2]. In recent years, the interest has been shifting to spontaneous facial expressions, as they are more ecologically valid. This is reflected by the recent publication of databases of spontaneous facial expressions [23], [36], [37], [38], [39], [40].

Due to their different characteristics, spontaneous expressions are more challenging to elicit and detect. These challenges include the detection of fast and subtle spontaneous expressions. Examples of spontaneous, fast, and subtle facial expressions are micro-expressions.

2.4 Micro-Expression Detection

Micro-expressions are brief, subtle, facial expressions that are leaked despite efforts to either deliberately or unconsciously conceal an emotion [41]. They can be considered as spontaneous, because they happen against the will of the person showing them, and better reflect the experienced affect.

Micro-expressions last only a fraction of a second. Because of their short duration, they are usually neither noticeable to an untrained eye nor to the people disclosing the expression. Besides the concealment, the main element in the definition is the duration. However, there seems to be a lack of consensus in their precise duration range [42], [43], [44].

Initially, the boundary between micro and macro-expressions was described as half a second by Ekman and Friesen [41], emphasizing that they are difficult to perceive for the untrained eye. However, in subsequent papers, he and other authors determined different boundaries. [44] described at least six different definitions, with an upper limit ranging from 1/25 of a second, to 1/2 second; and argued that the main difference between a macro and a micro expression is whether they can be perceived by an untrained eye. In their experiment, they determined that the critical point between distinguishing macro and micro expressions is about 1/5 seconds. Furthermore, they confirmed that a layperson's ability to perceive micro-expressions can be improved with training. In another attempt to determine an appropriate duration for micro-expressions, [43] elicited 109 fast expressions (less than 500 ms), and fitted distribution curves to their total and onset duration. Based on these, they conclude that a total duration of less than 500 ms plus an onset duration of less of 260 ms is suitable to define a micro-expression. Interestingly, they found out that these duration distributions seem to be related to the type of facial expression itself. Even though participants were following the same instructions to inhibit them, expressions of happiness usually lasted longer than expressions of disgust.

Previously micro-expressions have been detected mainly through computer vision methods [19], [24], [45], [46], [47], [48], [49]. Depending on the number of expressions identified, the results range from 54 percent accuracy in a leave-one-subject-out validation in [24] to 92 percent AUC in [49]. Again, the diversity in the methods used make it difficult to compare them. Additionally, it is recommended to use hi-speed cameras to record the data, because of the quick nature of such expressions [19], [23]. On the other hand, other sensors besides video cameras seem to be more robust in different contexts. EMG-based systems have the potential to detect such facial expressions because of their good temporal resolution. High sampling rates make it a promising tool to detect micro-expressions. Moreover, it is robust against head rotations and occlusion; and in a wearable, it could provide independence of movement. Despite this, to the best of our knowledge, micro-expression detection with EMG has not yet been explored. Thus, it remains a question whether these expressions can be identified with the same methods used to identify macro expressions from EMG.

Therefore, we propose to evaluate whether micro-expressions can be identified with an EMG-based wearable device. As a first step, this paper focuses on evaluating the feasibility of using distal surface EMG to detect micro-smiles. Micro-smiles were chosen because (1) micro-expressions are fast and subtle, and therefore represent a major challenge in facial expression recognition; and (2) smiles are related most of the time to positive affect, which is beneficial in the aforementioned application domains.

In the following sections, we will describe the proposed wearable device, the methods used to elicit micro-smiles, and the recognition algorithm. Furthermore, we argue for the convenience of this tool to annotate Ad, or any other video stimuli, and human-human social interactions. Finally, we discuss the results.

3 Wearable Device

The present work follows the design guidelines for facial expression detection using surface EMG provided by previous research [33], and extensively tested in various settings [34], [35], [50]. Fig. 1 shows the proposed arrangement. It uses four EMG channels placed on the sides of the face, on top of the *temporalis* and the *zygomaticus major* muscles [33]. Since distal EMG is measured, we do not identify the activity of each facial muscle, but a combination of their activities. Therefore, especial signal processing is required. The advantage of using distal EMG is that the electrodes do not

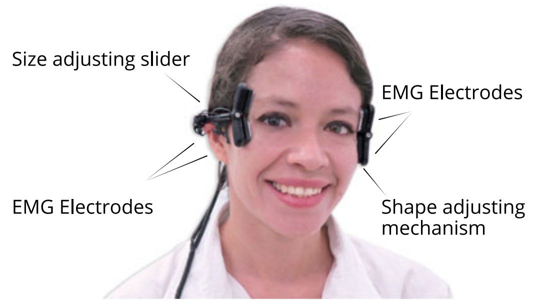

Fig. 1. The wearable used to record the micro-expressions. It includes ergonomic adjustment mechanisms. These allow to adjust four EMG electrode pairs to the head size and shape.

obtrude the muscle movement. Hence, the facial expressions of interest are not altered by wearing the device. Furthermore, its wearable nature allows for free movement, and good detection in spite of occlusion [35]. However, previous work applied such smile detection only to posed and/or macro-smiles. Micro-smile detection is more challenging due to the temporal and magnitude characteristics of these facial expressions. They are so fast and subtle that even humans experience difficulties perceiving them. As discussed in the following sections, we propose to improve the design of the wearable device; and to adapt the signal processing for the detection of micro-smiles. The proposed wearable approach is promising for real-time tracking of multiple people's micro-expressions. Usually CV methods are computationally expensive, and more often than not, are limited to the tracking of one person at a time.

3.1 Garment Design

The current system prototype consists of four surface EMG channels connected to a wireless transmitter. The position of the electrodes is on the sides of the face, on top of the temporalis and the zygomaticus major. Each channel consists of two active electrodes bonded together in a 20 by 10 mm box. This box is inserted in a placeholder, which in turn, is attached to a circlet with a bolt and screw. The purpose of the circlet is to keep the four channels on place (See Fig. 1). This is done by applying pressure on both sides of the face. The four placeholders can rotate slightly inwards, to adjust to the shape of the head of the wearer. Furthermore, the size of the circlet can be adjusted by screwing the attachment between the placeholders and the circlet.

3.2 Signal Processing

As a prototype of the functionality, the EMG signals picked up by the electrodes are transmitted to a laptop via Bluetooth, where they are analyzed using Matlab.

The surface EMG is recorded at 1 kHz sampling rate using a four channel Biolog DL-4000 system. 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. Next, the signals are decomposed in their Independent Components (IC), using Independent Component Analysis (ICA). The ICA allows to separate the distal EMG from different source muscles. Then the absolute value of the components is considered, and its Root-Mean Square value is calculated over overlapping windows of 100 ms, sliding one sample at a time. This was done to increase the temporal resolution of the algorithm, hence optimizing it for micro-smile detection. The aforementioned pre-processing was performed on all EMG data, for each participant. The resulting data is considered as input features to train a Neural Network (NN) with one hidden layer of four Sigmoid neurons (Fig. 2). Due to anatomical differences in muscle size and Body Mass, EMG is highly variable between subjects. Hence, within subject data was used to train the NN. Furthermore, given the limited availability of micro-smile samples, the no-expression data was under sampled to match the number of samples of the expression data [51], [52]. No-expression data was taken randomly from all the data. To validate this model, cross-validation with 70 percent train, 15 percent validation, and 15 percent test data is used. The neural network aims to compare micro-expressions with no-expression display. Micro-smiles are compared to no-expressions, as the electrode positions are optimized for positive expressions.

4 Experiment Design

The main purpose of this experiment is to assess the possibility of detecting fast and subtle smiles with distal EMG. To the best of our knowledge, no online database is available, which includes unobtrusive, distal EMG of micro-expressions, recorded in the proposed arrangement. Therefore, an experiment was designed for data collection. To elicit micro-expressions, a methodology similar to the one described in [23] was used. Furthermore, the video stimuli were mainly Ads that could be assessed with the proposed method. The experiment was within-subjects, where all participants watched all stimuli in a counterbalanced order.

4.1 Stimuli Selection

Due to the importance of positive affect cues as personal and product feedback, we decided to focus on eliciting smile expressions. Therefore, three Ad videos were selected from stimulus used in previous research [37]. Namely, "The force" (Video TF, 62 s), "House sitting" (Video HS, 30 s), "Parisian Love" (Video PL, 53 s). According to McDuff et al., using this stimuli they could collect more than 10,000 frames of smiles, hence we expected to get similar results. All videos were presented at 30 frames per second with 720 x 480 pixel resolution.

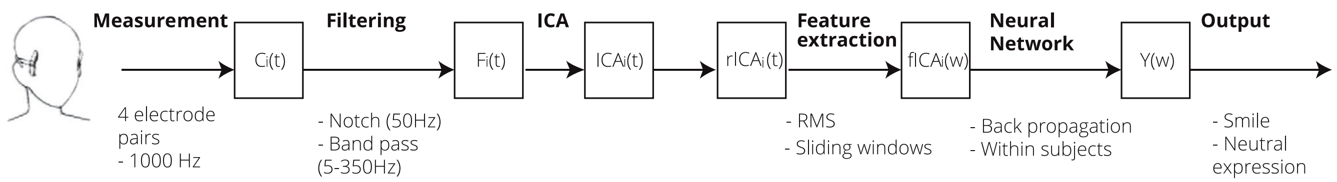

Fig. 2. Signal processing steps per participant.

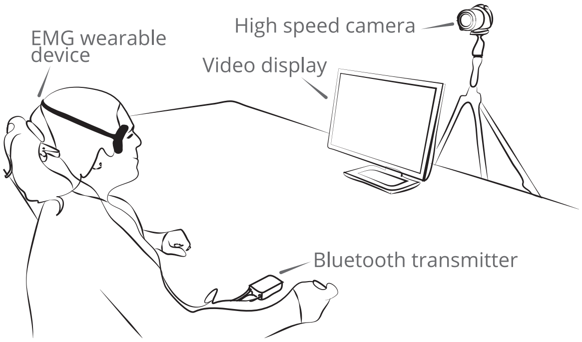

Fig. 3. Experiment setup. Participants wore the EMG device while sitting in front of a LCD screen where the stimuli videos were shown. The EMG data was transmitted via Bluetooth to a laptop. Finally, a camera was placed in front of them to record their facial expressions. This video was used as ground truth.

4.2 Pilot

One healthy participant (female, 34 years old) went through all the procedure as a pilot to evaluate the effect of the stimuli. Since the video PL did not elicit any facial expression, a fourth video was additionally included. The additional video was an edition of the 2011 Jimmy Kimmel Challenge "I Told My Kids I Ate All Their Halloween Candy" (Video HA, 2 min 9 s). As described in the following sections, the new video was more successful in provoking smiles.

4.3 Participants

Twenty-three voluntary participants took part on the study (average age = 26.9 years old, SD = 3.57). None of them had experience using the measuring device, and 14 participants had seen at least one of the videos before.

4.4 Procedure

Participants were provided with a general description of the test. The description stated that the purpose of the test was to rate several videos, while recording their facial EMG. Second, they provided their informed consent. Next, they were given the right to quit at any time. The task consisted in watching three videos in counterbalanced order. Before each video, they were asked to watch 30 seconds of black screen, as a baseline for the measurements. They were instructed to "keep a neutral face while watching the videos". Finally, the experimental setup is shown in Fig. 3.

4.5 Measurements

During the task, surface EMG and the participant's face were recorded simultaneously. The camera was a Canon Ivis HFg10 with HD resolution at 30 frames per second (fps) for the first five participants. For the last seven participants, the camera was changed to a Sony Cyber-shot DSC-RX100 II with 1920×1020 resolution at 120 fps. The purpose of this was to increase the number of frames in which the micro-smile would be shown, and therefore facilitate the job of the human coder. As mentioned before, a micro-expression can be argued to last as little as 1/25 of a second. Even with a normal camera recording at 30 fps, there would be at least one recorded frame of each micro-smile lasting 1/25 seconds. The number of frames is increased to four in the case of a 120 fps high-speed camera.

TABLE 1
Number of Expressions Elicited per Video

Stimuli	Macro-expressions		Micro-expressions		Total
	Smile	Other	Smile	Other	
1	28	14	3	14	60
2	18	18	6	11	53
3	3	18	0	10	31
4	157	69	23	28	277
Total	206	119	32	63	421

Video 1 lasted for 62 s, Video 2 for 30 s, Video 3 for 53 s, and Video 4 for 129 s. Micro-expressions are those lasting for less than 0.5 s. Other expressions refer to facial expressions that were not labeled as smiles.

5 Results

5.1 Video Coding

All the recordings of the participants' face were coded frame-by-frame for facial expressions by two experienced coders. High-frame rate videos were slowed down to facilitate the task. The labeling included coding for the onset, offset, and apex frames of the facial expression; the Facial Action Unit Systems Action Units that were present in the expression; and whether it was considered a smile or not, and laughter or not. The labeled AU were AU1, AU02, AU04, AU05, AU6, AU09, AU10, AU12, AU14, AU15, AU17, AU18, AU25, AU26, AU28, and AU38. Smiles were often a display of AU6 and/or AU12. However, the smile label was not assigned every time these AU occurred [37]. A facial expression was considered as each annotation of the appearance of an AU change from the face on resting state or a change from a different AU. The duration of the expression was calculated as the difference between the coded onset and the apex. All facial movements considered as swallowing, coughing, or sneezing were excluded.

A total of 421 facial expressions were identified by at least one human coder. These were displayed by 21 of the participants, two of them (Participant 4, 9) managed to keep a neutral face during all the videos.

Table 1 shows the number of expressions that were elicited by each video. From the elicited expressions, 238 were smiles; 177 were expressions faster than one second; 67 of the smiles were faster than one second; and 95 expressions were faster than 1/2 second, from which 32 were smiles. Expressions lasting less or equal than half a second were considered micro-expressions.

The Cohens Kappa Coefficient was used as a measure of inter-rater agreement [53]. For this paper, only the information of the duration of the expression plus the assessment of whether the expression was a smile or not were used. Therefore, the Kappa Coefficient was calculated on the frame-by-frame human coding on whether the participant was smiling or not. According to this, the Cohen's Kappa Coefficient was 0.4068 ($p < 0.01$).

5.2 EMG Signal Processing

Fig. 4 shows the EMG processing steps for participant seven, video four. Dark dotted squares indicate the smile labels given by the human coder. Four examples of the facial expressions are also shown. In here, only one out of four channels is shown. The gray signal is shows the raw

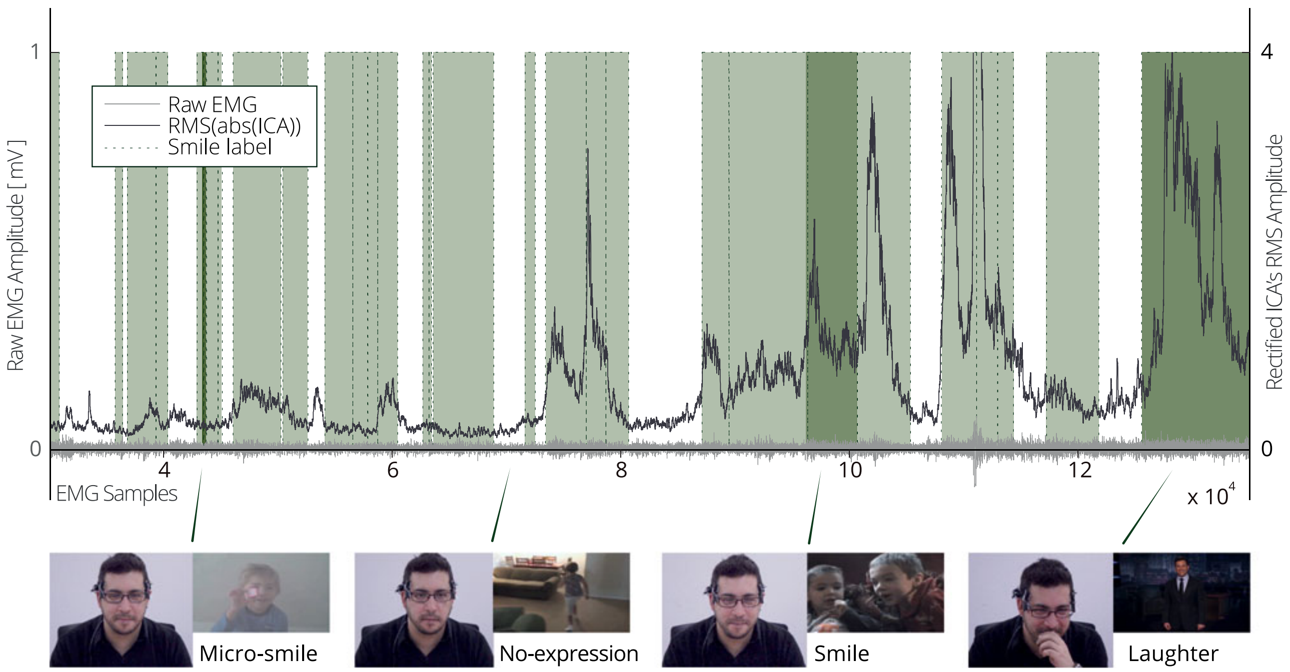

Fig. 4. Processing stages for participant 7. The dark dotted squares indicate the smile labels given by a human coder. An example frame shot of the coded expression is given below the plot. The stimuli video frame seen at that moment is also shown. The gray signal is one of the four raw EMG channels, and its magnitude is shown in the left Y axis. The black signal corresponds to the RMS of one of the four rectified ICA components. Its magnitude is shown in the right Y axis. The latter was used as input for the Neural Network. Axis X represents the EMG sample number, as synchronized with the video. The participant starts with moderate smiles, sometimes micro-smiles. The affect builds up, until he breaks in laughter by the end of the video.

EMG, and its magnitude is shown in the left Y axis. The black signal corresponds to the RMS of a rectified ICA component. Its magnitude is shown in the right Y axis. The differences can be seen in feature magnitude for a non-expressive face, a micro-smile, and macro-smiles. Although the plot does not show the activity from the start of the video, it can clearly be observed that the displayed smiles are faster and subtler earlier in the video. By the end of the video, the subject cannot contain the laughter anymore. Finally, it was observed that five participants tended to cover they mouth to hide their faces when they cannot contain laughter. Other coping strategies were masking the expressions with other facial movements like swallowing and wetting their lips.

Sixteen out of 23 participants displayed micro-expressions, and 10 of them displayed micro-smiles. Only their data was used for further analysis. Table 2a shows the precision and recall achieved for micro-smile detection for each participant, for the neural network results between micro-smiles and no-expression. Table 2b shows similar results but for classification of smiles lasting at most one second. Interestingly, the achieved performance seems to decrease as longer smiles are included in the data set.

6 Application Case Study

EMG-based facial expression detection has the potential of being used to annotate the interaction between Ads or other stimuli, and human-human interactions. Previously, [37] showed that smiles detected using CV over the internet can be used to predict online media effectiveness. However, this study was limited to macro-smiles of a single viewer. As shown in previous sections, EMG has critical advantages in detecting micro-smiles in real time, and where occlusion is likely to occur. In this section, we show an exploratory study on the performance of simultaneous detection of multiple people's fast and subtle smile detection using our proposed method.

7 Experiment Design

The experiment design consisted of two blocks. The first block, from now on called “Conditioned Block”, was counterbalanced with a second block, from now on referred to as “Free Block”.

The Conditioned Block was identical to the experiment setup of the previous experiment, except that two participants were invited to watch the videos simultaneously. In the Free Block, the same two participants watched four new videos, and they were invited to comment and relax, as if they were at home watching the videos with a friend. Besides counterbalancing the order of the blocks, the order of the four videos within each block was counterbalanced as well.

7.1 Participants

Eight voluntary participants took part on the study (four female, average age = 29.25 years old, SD = 1.71). None of them had experience using the measuring device. All participants had seen at least one of the videos before.

7.2 Stimuli

For the micro-smile elicitation block, the same videos from the previous experiment were used. Additionally, three new Ad videos were selected, plus a video showing funny and cute kid behavior. The videos are, “Baby expectancy“ (Video 5, 29 s), “Fun kiddies” (Video 6, 2 min 18 s),

TABLE 2
Smile Detection Results

(a) Classification results of micro-smiles (lasting less than 0.5 s) and neutral faces. RMS was used as feature.

Participant	Micro-smile detection			Number of smile EMG samples	Number of smiles	Age	Gender
	Precision	Recall	Accuracy				
1	NA	NA	NA	0	0	25	Masculine
2	100%	100%	100%	434	1	26	Feminine
3	NA	NA	NA	0	0	27	Masculine
4	NA	NA	NA	0	0	24	Masculine
5	NA	NA	NA	0	0	35	Masculine
6	100%	100%	100%	418	1	26	Masculine
7	98.1%	98.4%	98.2%	1,855	7	34	Masculine
8	NA	NA	NA	0	0	34	Masculine
9	NA	NA	NA	0	0	29	Feminine
10	99.8%	99.9%	99.8%	2,318	10	28	Masculine
11	NA	NA	NA	0	0	28	Feminine
12	NA	NA	NA	0	0	28	Masculine
13	79.3%	82.4%	81.20%	1,160	3	27	Feminine
14	100%	100%	100%	243	1	24	Masculine
15	100%	100%	100%	434	2	26	Masculine
16	94.6%	96.8%	95.8%	1,605	4	23	Masculine
17	NA	NA	NA	0	0	31	Feminine
18	NA	NA	NA	0	0	24	Feminine
19	100%	100%	100%	401	1	23	Masculine
20	100%	100%	100%	286	2	22	Masculine
21	NA	NA	NA	0	0	24	Masculine
22	NA	NA	NA	0	0	25	Masculine
23	NA	NA	NA	0	0	25	Feminine
Total				9,113	32		

(b) Classification results of smiles lasting less than a second and neutral faces. RMS was used as feature.

Participant	Micro-smile detection			Number of smile EMG samples	Number of smiles
	Precision	Recall	Accuracy		
1	94.9%	96.9%	95.9%	801	1
2	96.3%	99.6%	98.00%	1,167	2
3	NA	NA	NA	0	0
4	NA	NA	NA	0	0
5	NA	NA	NA	0	0
6	79.8%	88.1%	84.5%	3,705	5
7	89.5%	94.9%	92.3%	8,356	17
8	98.5%	93.5%	95.8%	926	1
9	NA	NA	NA	0	0
10	82.8%	91.7%	87.6%	5,041	14
11	100%	100%	100%	2,045	3
12	NA	NA	NA	0	0
13	99.9%	99.5%	99.7%	3,940	4
14	100%	100%	100%	1,128	2
15	100%	100%	100%	393	2
16	64.7%	80.3%	74.4%	5,895	11
17	83.1%	81.1%	81.9%	651	1
18	NA	NA	NA	0	0
19	100%	99.9%	99.9%	2,754	2
20	100%	100%	100%	286	2
21	NA	NA	NA	0	0
22	NA	NA	NA	0	0
23	NA	NA	NA	0	0
Total				37,088	67

RMS was used as feature. The number of smiles represent the amount of smiles observed. The number of smile EMG samples represents the total amount of EMG samples taken at 1 kHz from onset till offset of all reported smile facial expressions.

"Brotherhood" (Video 7, 53 s), "Dirt Devil" (Video 8, 1 min 28 s). All videos were presented to the participants at 30 frames per second with 720 × 480 pixel resolution.

7.3 Measurements

During the task, surface EMG and the face of both participants were recorded simultaneously. The camera used was

TABLE 3
Number of Expressions Elicited per Video

Stimuli	Macro-expressions		Micro-expressions		Total
	Smile	Other	Smile	Other	
1	10	5	4	0	19
2	6	9	0	5	20
3	0	5	0	5	10
4	15	5	0	7	27
5	13	18	3	10	44
6	59	48	14	12	133
7	17	11	1	4	33
8	22	14	2	9	47
Total	142	115	24	52	333

Video 1 lasted for 62 s, Video 2 for 30 s, Video 3 for 53 s, Video 4 for 129 s, Video 5 for 30 s, Video 6 for 158 s, Video 7 for 57 s, and Video 8 for 88 s. Micro-expressions are those lasting for less than 0.5 s. Other expressions refer to facial expressions that were not labeled as smiles.

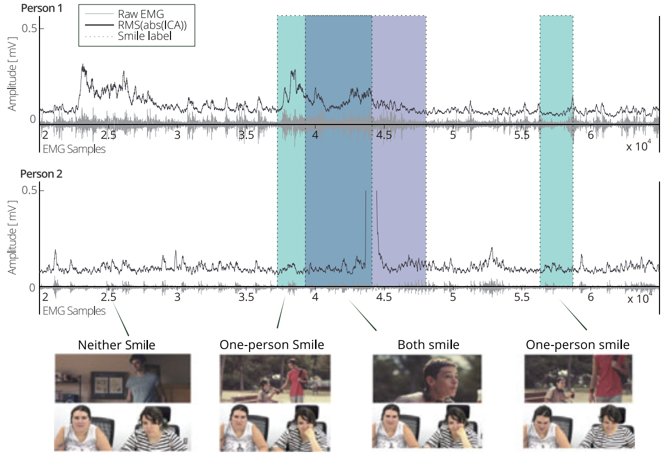

Fig. 5. The wearable can also be used to measure multiple users at the same time, a hardware trigger was used to synchronize multiple devices and the stimuli presentation. With this method, it can be observed when people smile simultaneously or alone.

a Sony Cyber-shot DSC-RX10 II with 1920 × 1020 resolution at 120 fps. For device-device-stimuli synchronization purposes, a hardware trigger was designed. This consisted of a micro-controller interfacing via serial port with a laptop used to present the stimuli and send the triggers to the EMG acquisition device.

8 Analysis and Results

All the recordings of the participants' face were coded frame-by-frame for facial expressions by one experienced coder, similarly as in the previous study. From the elicited expressions, 166 were smiles, 76 were expressions faster than half a second, and 24 were smiles faster than half a second (Table 3). In this study, six out of eight participants tended to rotate their heads to face the other participant, to cover their faces depending on the situation and to talk quite often. Fig. 5 shows the results of human coding for totally and partially covering the face, looking towards each other, and looking away to the camera. Smile labels are also shown in the figure. Table 4 summarizes the results of the classification with the proposed method for micro-smiles

TABLE 4
Classification Results

(a) Classification results of micro-smiles (lasting less than 0.5s) and neutral faces.

Participant	Spontaneous smile detection			Number expression samples	Number expressions	Age	Gender
	Precision	Recall	Accuracy				
1	NA	NA	NA	0	0	29	Masculine
2	NA	NA	NA	0	0	31	Masculine
3	NA	NA	NA	0	0	30	Feminine
4	96.8%	97.5%	97.2%	493	1	29	Feminine
5	100%	100%	100%	1,120	4	30	Feminine
6	NA	NA	NA	0	0	30	Masculine
7	88.8%	91.3%	90.2%	3,686	17	25	Feminine
8	100%	100%	100%	886	2	30	Masculine
Total				6,185	24		

(b) Classification results of smiles lasting less than two seconds and neutral faces.

Participant	Spontaneous smile detection			Number expression samples	Number expressions
	Precision	Recall	Accuracy		
1	NA	NA	NA	0	0
2	97.7%	99.0%	98.4%	5,529	4
3	96.1%	95.1%	95.6%	2,619	2
4	87.4%	93.3%	90.6%	3,353	3
5	86.3%	87.2%	86.8%	58,050	28
6	NA	NA	NA	0	0
7	79.3%	84.8%	82.6%	19,074	39
8	85.1%	92.3%	89.0%	2,521	7
Total				91,146	83

RMS was used as feature. The number of smiles represent the amount of smiles observed. The number of smile EMG samples represents the total amount of EMG samples taken at 1 kHz from onset till offset of all reported smile facial expressions.

and for longer-lasting smiles. Here, 2 s or less was chosen arbitrarily to exemplify how performance varies with the duration. Longer smiles and laughter present more variation in the EMG due to muscle fatigue or multiple EMG bursts, respectively. Therefore, they are more difficult to identify as a unit. Future work should explore possibilities to address these challenges.

9 Discussion

This paper explored the feasibility of detecting micro-smiles using surface EMG. The results showed that micro-smiles can be distinguished from a neutral face using EMG, with good accuracy. Therefore, EMG pre-processing and classification methods seem to be also useful to analyze micro-smile expressions.

The achieved accuracy to identify micro-smiles was very good. This high accuracy cannot be explained due to overfitting, as we made sure to apply proper cross-validation. Nevertheless, an important challenge is the unbalanced nature of the data. Micro-smiles are very rare compared to the no-expression class. To address this issue, we under sampled the non-expression data. The results shown in Table 2a are reporting the accuracy for this case. Interestingly, the same algorithm yielded less accurate results if the inclusion criteria was extended to fast and subtle smiles lasting less than a second (Table 2b). In the latter case, accuracy ranged from 74 till 100 percent depending on the participant. This could be explained by the duration of the expressions themselves. As the duration time is shorter, there are less variations in the EMG intensity as can be observed in longer expressions. In longer smiles, the relationship between detected electrical output at an EMG site, and the mechanical force exerted by a muscle may change over time as a function of fatigue [30]. This variability would make it more difficult for the machine learning algorithm to define a clear boundary between samples of different expressions. Furthermore, previous work has suggested that muscle contractions in spontaneous expressions peak simultaneously [2], which could contribute to the success of the EMG-based classification. An alternative explanation could be that usually smiles happen on top of other smiles. It was observed that once some participant smiles, the smile tends to last for a long period of time, and after a while, this smile is further extended in reaction to a funnier stimuli event. This could also contribute to the clear magnitude difference between the neutral expression and a micro-smile.

Furthermore, the temporal resolution and portability of this device would allow to provide real-time feedback, if desired. This can be used for quantification applications of positive facial expressions. In spite of using a 100 ms window for processing, the temporal resolution is still 1 kHz, because the window is sliding every sample. However, there is still some work to be done before bringing this wearable device to a real-time application setting. Only the results of an offline setup are showed here. For real-time applications, the micro-smile recognition should be ported to an online classification system. Moreover, one of the challenges would be the calibration of the system for each individual. Previous work on CV shows that inter-subject training and validation leads to poor results [24]. Regarding EMG, high inter-subject and inter-session variability is also expected. This is mainly due to anatomical differences in muscle size and position, and in Body Mass Index. In between sessions, one of the main sources of error are the differences in electrode position [33], [54], and changes in skin conductance; suggesting that individual calibration is preferred to ensure good performance. However, elicitation of such fast and subtle spontaneous expressions for calibration purposes is a challenging task.

Previously, Yan et al. [55] discussed that the stronger the emotion felt, the more micro-expressions could be elicited. Furthermore, a high stakes situation is often required to elicit micro-expressions. In our case, the videos were short, they were not watched for the first time, and they elicited mild emotional content. Furthermore, the intensity of the facial expressions has also been argued to depend on the elicitation paradigm used. The neutralization paradigm implies that facial expressions are inhibited with strong intent, and therefore not leaked easily. On the other hand, during our tests, we could observe the two extremes: participants 4 and 9 did not leak any expression, whereas most of them were showing wide smiles. Other participants broke in laughter in several occasions. This resulted in all their smiles lasting longer than half a second. This phenomenon was observed before by [55]. They reported that expressions of happiness tend to last longer than the micro-expression threshold, and that micro-expressions of disgust are much more frequent. In our data, 34.3 percent of the elicited micro-expressions were smiles, despite the stimuli being rated as positive, and the smiles being 63.5 percent of the macro-expressions. Even though with the proposed experimental setup, only about 44 percent of the participants displayed micro-smiles. This is not unexpected. In their experiment Yan et al. also discussed that "As for micro-expressions of happiness, one may feel surprised as to why so few were elicited. Though happiness feelings were easily elicited when watching amusing video episodes, these elicited smiles or laughter bursts are lasting facial expressions and do not fit the criteria of micro-expression. Thus, most of the elicited happy facial expressions were categorized as conventional facial expressions." The phenomena we observed is similar, despite using different stimuli. Furthermore, when the stimuli are strong enough, the displayed affective expression is often long-lasting laughter. It requires quite some effort to conceal laughter, and while some participants are very good in neutralizing their facial expressions, most others are not. Therefore, we argue that a fast and subtle smile is a first laughter burst that is quickly contained, and that the proposed method and device is thus able to detect both.

Despite the number of micro-smiles might seem small, the high sampling frequency of EMG allows to obtain 500 samples (at most) per micro-smile. In other words, we could analyze about 9,113 EMG samples of micro-smiles in the single-participant experiment, and 6,185 in the pairs-experiment.

Even though this was the first effort to evaluate the detection of fast and subtle spontaneous facial expressions using an EMG-based wearable device, we only tested with smiles. Future work should address the possibility of extending the detection to more facial expressions. Smiles were a good first step because of their potential in different application domains, especially as a measure of wellbeing; as measure of acceptation of a product or therapy; or as a positive reward in learning applications.

The current wearable device electrode positions were selected based in [33], and are optimized for smile detection. This was done in the aforementioned research by considering a multi-attribute decision making process, facial morphology, and EMG amplitude. Moreover, the device is able to detect other expressions such as posed frowning. Including more electrodes would be helpful to expand the wearables spatial resolution to detect more facial expressions. In the future, the use of dense and compact grids of electrodes around the face [54] seem a good alternative to achieve so without covering it. Covering the face is an undesired situation as it makes the users aware of their own facial expressions, and limits the movement of the skin [30], [33].

Although CV-based micro-expression methods have better spatial resolution, state-of-the-art algorithms are still sensitive to occlusion; computationally expensive; difficult to implement in a real-time feedback setting; and often get heavier when there is more than one face on scene, causing a less accurate detection. EMG poses a good alternative to robust micro-expression detection, and a potential replacement to human video coding. Human perception of micro-expressions requires training, and video coding of these is cumbersome and time-consuming. EMG provides accurate automatic detection, as it profits from complementary information to what the human cannot see.

Finally, EMG-based wearables can provide event-related smile detection. Identifying the relationship between a stimuli event and (micro-) smiles can better provide information about the synchronization or de-synchronization of the smiles between several humans and/or a stimulus. This automatic multiple-user facial expression annotation can support experts in other domains to identify salient elements in their Ads, products, or therapies. This paper showed the feasibility to annotate a stimulus with positive affect cues, even if they are as fast and subtle as micro-smiles.

10 Conclusion and Future Work

We presented a wearable device which can provide laughter analysis in human-human communication. Although laughter is characterized by complex expressive behavior, in particular, we focus on the dynamics of facial expressions of positive affect. We analyzed fast and subtle smiles at the level of micro-expressions, and showed a method to use the detection to annotate stimuli. In this manner, progression from short smiles to laughter can be observed along with the participants' experience. We argued for the advantages of using a wearable approach for such detection, as the computer vision approach has some major drawbacks such as inaccurate detection in cases of (1) occlusion; (2) face-to-face human-human interaction; and (3) computational expensiveness of micro-expression detection. In this paper, we focused on (3) and showed an example of (2). We made the first effort to prove the feasibility of detecting micro-smiles with a wearable device. We believe this is an important first step for automatic analysis of spontaneous smiles and laughter in human-human communication. As observed from our results, in ecologically valid settings people tend to accompany laughter with head and hand movements. Therefore, other major expressive modalities such as speech, body movements, and postural attitudes, might be complementary to annotate the situation and eventually infer its meaning. In the future, we plan to integrate a multimodal detection in our wearable approach.

Acknowledgments

The authors would like to thank Dushyantha Jayatilake, and Airi Tsuji for their contribution in the EMG circlet design and video coding, respectively. Furthermore, we would like to thank the people who participated in the studies.

References

[1] T. D. Wilson and D. T. Gilbert, "Affective forecasting," *Advances Exp. Social Psychology*, vol. 35, pp. 345–411, 2003.

[2] P. Ekman and E. Rosenberg, *What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS)*, 2nd ed., P. Ekman and E. Rosenberg, Eds. London, U.K.: Oxford Univ. Press, 2005.

[3] C. M. Dillon and J. E. Carr, "Assessing indices of happiness and unhappiness in individuals with developmental disabilities: A review.pdf," *Behavioral Interventions*, vol. 22, pp. 229–244, 2007.

[4] A. Thieme, J. Wallace, T. D. Meyer, and P. Olivier, "Designing for mental wellbeing: Towards a more holistic approach in the treatment and prevention of mental illness," in *Proc. Brit. HCI Conf.*, 2015, pp. 1–10. [Online]. Available: http://doi.acm.org/10.1145/2783446.2783586

[5] S. Matsuda and J. Yamamoto, "Research in autism apectrum disorders intervention for increasing the comprehension of affective prosody in children with autism spectrum disorders," *Res. Autism Spectr. Disorders*, vol. 7, no. 8, pp. 938–946, 2013. [Online]. Available: http://dx.doi.org/10.1016/j.rasd.2013.04.001

[6] J. Cockburn, M. Bartlett, J. Tanaka, J. Movellan, M. Pierce, and R. Schultz, "SmileMaze: A tutoring system in real-time facial expression perception and production in children with autism spectrum disorder," in *Proc. IEEE Int. Conf. Automat. Face Gesture Recognit.*, 2008, pp. 978–986.

[7] C. Breazeal and R. Brooks, "Robot emotion: A functional perspective," in *Who Needs Emotions: The Brain Meets the Robot*, J.-M. Fellous and M. A. Arbib, Eds. Cambridge, MA, USA: MIT Press, 2005, pp. 271–310.

[8] J. Laparra-Hernandez, J. Belda-Lois, E. Medina, N. Campos, and R. Poveda, "EMG and GSR signals for evaluating user's perception of different types of ceramic flooring," *Int. J. Ind. Ergonom.*, vol. 39, pp. 326–332, 2009.

[9] D. Mcduff, R. Kaliouby, D. Demirdjian, and R. Picard, "Predicting online media effectiveness based on smile responses gathered over the internet," in *Proc. 10th IEEE Int. Conf. Workshops Autom. Face Gesture Recognit.*, 2013, pp. 1–7.

[10] M. Soleymani, S. Asghari-esfeden, Y. Fu, and M. Pantic, "Analysis of EEG signals and facial expressions for continuous emotion detection," *IEEE Trans. Affect. Comput.*, vol. 7, no. 1, pp. 17–28, Jan.–Mar. 2016.

[11] M. B. Harms, A. Martin, and G. L. Wallace, "Facial emotion recognition in autism spectrum disorders: A review of behavioral and neuroimaging studies," *Neuropsychology Rev.*, vol. 20, no. 3, pp. 290–322, 2010. [Online]. Available: http://link.springer.com/10.1007/s11065–010-9138-6

[12] L. M. Oberman, P. Winkielman, and V. S. Ramachandran, "Slow echo : Facial EMG evidence for the delay of spontaneous, but not voluntary, emotional mimicry in children with autism spectrum disorders," *Develop. Sci.*, vol. 4, pp. 510–520, 2009.

[13] D. N. Mcintosh, A. Reichmann-decker, P. Winkielman, and J. L. Wilbarger, "When the social mirror breaks : Deficits in automatic, but not voluntary, mimicry of emotional facial expressions in autism," *Develop. Sci.*, vol. 3, no. 9, pp. 295–302, 2006.

[14] D. Galati, K. R. Scherer, and P. E. Ricci-Bitti, "Voluntary facial expression of emotion: Comparing congenitally blind with normally sighted encoders." *J. Personality Social Psychology*, vol. 73, no. 6, pp. 1363–79, Dec. 1997. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/9453990

[15] W. Ruch and P. Ekman, "The expressive pattern of laughter," in *Emotion Qualia, and Conciousness*, A. Kaszniak, Ed. Tokyo, Japan: World Scientific Publisher, 2001, pp. 426–443. [Online]. Available: https://www.paulekman.com/wp-content/uploads/2013/07/The-Expressive-Pattern-Of-Laughter.pdf

[16] S. Cosentino, S. Sessa, and A. Takanishi, "Quantitative laughter detection, measurement, and classification—A critical survey," *IEEE Rev. Biomed. Eng.*, vol. 9, pp. 148–162, 2016. [Online]. Available: http://ieeexplore.ieee.org/document/7403873/

[17] V. Bettadapura, "Face expression recognition and analysis: The state of the art," *CoRR*, 2012. [Online]. Available: http://arxiv.org/abs/1203.6722

[18] S. L. Happy and A. Routray, "Automatic facial expression recognition using features of salient facial patches," *IEEE Trans. Affect. Comput.*, vol. 6, no. 1, pp. 1–12, Jan.–Mar. 2015.

[19] S. Polikovsky, Y. Kameda, and Y. Ohta, "Facial micro-expression detection in hi-speed video based on facial action coding system (FACS)," *IEICE Trans. Inf. Syst.*, vol. E96-D, no. 1, pp. 81–92, 2013.

[20] B. Fasel and J. Luettin, "Automatic facial expression analysis : A survey," *Pattern Recognit.*, vol. 36, pp. 259–275, 2003.

[21] J. H. Janssen, et al., "Machines outperform laypersons in recognizing emotions elicited by autobiographical recollection," *Human-Comput. Interaction*, vol. 28, no. 6, pp. 479–517, 2013.

[22] P. Ekman and W. P. Friesen, "Measuring facial movement with the facial action coding system," in *Emotion in the Human Face*, P. Ekman, Ed., 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 1982, ch. 9, pp. 178–211.

[23] W. J. Yan, et al., "CASME II: An improved spontaneous micro-expression database and the baseline evaluation," *PLoS One*, vol. 9, no. 1, pp. 1–8, 2014.

[24] X. Li, T. Pfister, X. Huang, G. Zhao, and M. Pietikainen, "A spontaneous micro-expression database: Inducement, collection and baseline," in *Proc. 10th IEEE Int. Conf. Workshops Autom. Face Gesture Recognit.*, 2013, pp. 1–6.

[25] R. Gross, J. Shi, and J. Cohn, "Quo vadis face recognition?" in *Proc. 3rd Workshop Empirical Eval. Methods Comput. Vis.*, 2001, pp. 119–132.

[26] G. Zhao, X. Huang, M. Taini, S. Z. Li, and M. Pietikäinen, "Facial expression recognition from near-infrared videos," *Image Vis. Comput.*, vol. 29, no. 9, pp. 607–619, Aug. 2011. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0262885611000515

[27] T. Pfister, X. Li, G. Zhao, and M. Pietikäinen, "Differentiating spontaneous from posed facial expressions within a generic facial expression recognition framework," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops*, 2011, pp. 868–875.

[28] G. E. Schwartz, P. L. Fair, P. Salt, M. R. Mandel, and G. L. Klerman, "Facial expression and imagery in depression: An electromyographic study," *Psychosomatic Med.*, vol. 38, no. 5, pp. 337–47, 1976. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/981492

[29] M. Balconi, G. Lecci, and V. Trapletti, "What do facial expressions of emotion express in young children? The relationship between facial display and EMG measures," *Neurophychological Trends*, vol. 15, pp. 7–23, 2014.

[30] D. Matsumoto, P. Ekman, and A. Fridlund, "Analyzing nonverbal behavior," in *Practical Guide to Using Video in the Behavioral Sciences*, P. W. Dowrick, Ed. Hoboken, NJ, USA: Wiley, 1991, ch. 10, pp. 153–165.

[31] M. Hamedi, S.-H. Salleh, M. Astaraki, and A. M. Noor, "EMG-based facial gesture recognition through versatile elliptic basis function neural network," *Biomed. Eng. Online*, vol. 12, no. 1, 2013, Art. no. 73.

[32] Y. Chen, Z. Yang, and J. Wang, "Eyebrow emotional expression recognition using surface EMG signals," *Neurocomput.*, vol. 168, pp. 871–879, 2015. [Online]. Available: http://dx.doi.org/10.1016/j.neucom.2015.05.037

[33] A. Gruebler and K. Suzuki, "Design of a wearable device for reading positive expressions from facial EMG signals," *IEEE Trans. Affect. Comput.*, vol. 5, no. 3, pp. 227–237, Jul.–Sep. 2014.

[34] Y. Takano and K. Suzuki, "Affective communication aid using wearable devices based on biosignals," in *Proc. Conf. Interaction Des. Children*, 2014, pp. 213–216. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2593968.2610455

[35] 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," *J. Autism Develop. Disorders*, vol. 44, no. 3, pp. 685–693, 2014.

[36] S. Koelstra, et al., "DEAP: A database for emotion analysis ;using physiological signals," *IEEE Trans. Affect. Comput.*, vol. 3, no. 1, pp. 18–31, Jan.–Mar. 2012.

[37] 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'," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Workshops*, 2013, pp. 881–888.

[38] S. M. Mavadati, M. H. Mahoor, K. Bartlett, P. Trinh, and J. F. Cohn, "DISFA: A spontaneous facial action intensity database," *IEEE Trans. Affect. Comput.*, vol. 4, no. 2, pp. 151–160, Apr.–Jun. 2013.

[39] S. Wang, et al., "Analyses of a multimodal spontaneous facial expression database," *IEEE Trans. Affect. Comput.*, vol. 4, no. 1, pp. 34–46, Jan.–Mar. 2013.

[40] M. K. Abadi, R. Subramanian, S. M. Kia, P. Avesani, I. Patras, and N. Sebe, "DECAF: MEG-based multimodal database for decoding affective physiological responses," *IEEE Trans. Affect. Comput.*, vol. 6, no. 3, pp. 209–222, Jul.–Sep. 2015.

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

[42] S. Porter and L. Ten Brinke, "Reading between the lies," *Psychological Sci.*, vol. 19, no. 5, pp. 508–514, 2008. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/18466413

[43] W.-J. Yan, W. Qi, L. Jing, Y.-H. Chen, and X. Fu, "How fast are the leaked facial expressions : The duration of micro-expressions," *J. Nonverbal Behavior*, vol. 37, no. 4, pp. 217–230, 2013. [Online]. Available: http://link.springer.com/10.1007/s10919–013-0159-8

[44] X.-B. Shen, Q. Wu, and X.-L. Fu, "Effects of the duration of expressions on the recognition of microexpressions," *J. Zhejiang Univ. Sci. B (Biomed. Biotechnol.)*, vol. 13, no. 3, pp. 221–230, 2012. [Online]. Available: http://www.springerlink.com/index/10.1631/jzus.B1100063

[45] M. Shreve, S. Godavarthy, D. Goldgof, and S. Sarkar, "Macro- and micro-expression spotting in long videos using spatio-temporal strain," in *Proc. IEEE Int. Conf. Autom. Face Gesture Recognit. Workshops*, 2011, pp. 51–56.

[46] X. Ben, P. Zhang, R. Yan, M. Yang, and G. Ge, "Gait recognition and micro-expression recognition based on maximum margin projection with tensor representation," *Neural Comput. Appl.*, 2015. [Online]. Available: http://dx.doi.org/10.1007/s00521–015-2031-8

[47] F. Xu, J. Zhang, and J. Wang, "Microexpression identification and categorization using a facial dynamics map," *IEEE Trans. Affect. Comput.*, vol. 8, no. 2, pp. 254–267, Apr.–Jun. 2017.

[48] X. Huang, G. Zhao, X. Hong, W. Zheng, and M. Pietikäinen, "Spontaneous facial micro-expression analysis using spatiotemporal completed local quantized patterns," *Neurocomput.*, vol. 175, pp. 564–578, 2015. [Online]. Available: http://dx.doi.org/10.1016/j.neucom.2015.10.096

[49] Z. Xia, X. Feng, J. Peng, X. Peng, and G. Zhao, "Spontaneous micro-expression spotting via geometric deformation modeling," *Comput. Vis. Image Understanding*, vol. 147, pp. 87–94, 2015. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S1077314215002702

[50] M. Hirokawa, A. Funahashi, Y. Itoh, and K. Suzuki, "A doll-type interface for real-time humanoid teleoperation in robot-assisted activity," in *Proc. IEEE Int. Symp. Robot Human Interactive Commun.*, 2014, pp. 174–175.

[51] H. He and E. A. Garcia, "Learning from imbalanced data," *IEEE Trans. Knowl. Data Eng.*, vol. 21, no. 9, pp. 1263–1284, Sep. 2009.

[52] L. A. Jeni, J. F. Cohn, and F. De La Torre, "Facing imbalanced data recommendations for the use of performance metrics," in *Proc. Int. Conf. Affect. Comput. Intell. Interaction Workshops*, 2013, pp. 245–251.

[53] J. Cohen, "Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit," *Psychological Bulletin*, vol. 70, no. 4, pp. 213–220, 1968. [Online]. Available: http://content.apa.org/journals/bul/70/4/213

[54] C. Amma, T. Krings, J. Böer, and T. Schultz, "Advancing muscle-computer interfaces with high-density electromyography," in *Proc. 33rd Annu. ACM Conf. Human Factors Comput. Syst.*, 2015, pp. 929–938. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2702123.2702501

[55] W.-J. Yan, Q. Wu, J. Liang, Y.-H. Chen, and X. Fu, "How fast are the leaked facial expressions: The duration of micro-expressions," *J. Nonverbal Behavior*, vol. 37, no. 4, pp. 217–230, 2013. [Online]. Available: http://link.springer.com/10.1007/s10919-013-0159-8

Monica Perusquía-Hernández received the BSc degree in electronic systems engineering from the Instituto Tecnolgico y de Estudios Superiores de Monterrey, Mexico, in 2009 and the MSc degree in human-technology interaction and the professional doctorate in engineering in user-system interaction from the Eindhoven University of Technology, the Netherlands, in 2012 and 2014, respectively. Currently she is working toward the PhD degree at the Empowerment Informatics Program, University of Tsukuba, Japan. Her research interests include affective computing, biosignal processing, augmented human technology, and artificial intelligence. She is a student member of the IEEE.

Mazakasu Hirokawa received the engineer's degree from the College of Engineering Systems, University of Tsukuba, Japan, in 2008, and the MEng and PhD degrees from the Graduate School of Systems and Information Engineering, University of Tsukuba, in 2011 and 2014, respectively. He is currently assistant professor at the Empowerment Informatics Program, University of Tsukuba, Japan. He is a member of the IEEE.

Kenji Suzuki received the BS degree in physics, and the ME and PhD degrees in pure and applied physics from Waseda University, Japan, in 1997, 2000 and 2003, respectively. He is currently a full professor in the Center for Cybernics Research, and principal investigator of Artificial Intelligence Laboratory, University of Tsukuba, Japan. He was also a visiting researcher at the Laboratory of Physiology of Perception and Action, College de France in Paris, and the Laboratory of Musical Information, University of Genoa, Italy. His research interests include augmented human technology, assistive and social robotics, humanoid robotics, biosignal processing, social playware, and affective computing. He is a member of the IEEE.