The Invisible Potential of Facial Electromyography

A Comparison of EMG and Computer Vision when Distinguishing Posed from Spontaneous Smiles

Monica Perusquía-Hernández
NTT Communication Science Laboratories
perusquia@ieee.org

Saho Ayabe-Kanamura
Faculty of Human Sciences, University of Tsukuba
sahoaya@human.tsukuba.ac.jp

Kenji Suzuki
Artificial Intelligence Laboratory, University of Tsukuba
kenji@ieee.org

Shiro Kumano
NTT Communication Science Laboratories
kumano.shiro@lab.ntt.co.jp

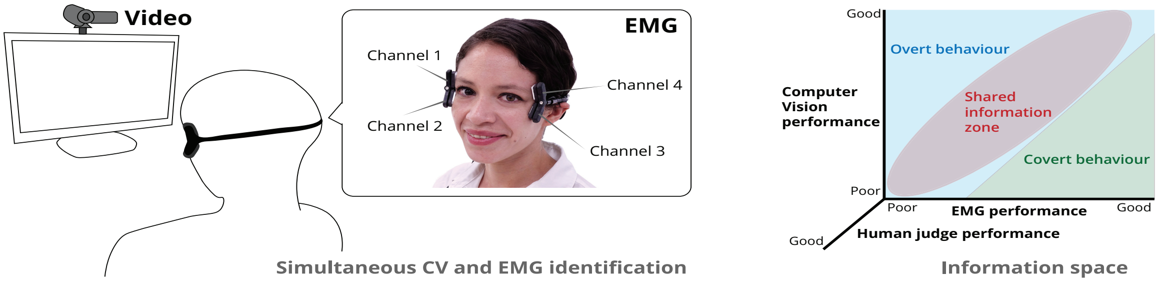

Figure 1: Wearable facial electromyography (EMG) measured distally has enabled the simultaneous EMG and facial video recording without electrodes obstructing the face. This study investigated the extent to which covert behavior measured by EMG contributes to the identification of posed and spontaneous smiles, and compared it with computer vision (CV) identification.

ABSTRACT

Positive experiences are a success metric in product and service design. Quantifying smiles is a method of assessing them continuously. Smiles are usually a cue of positive affect, but they can also be fabricated voluntarily. Automatic detection is a promising complement to human perception in terms of identifying the differences between smile types. Computer vision (CV) and facial distal electromyography (EMG) have been proven successful in this task. This is the first study to use a wearable EMG that does not obstruct the face to compare the performance of CV and EMG measurements in the task of distinguishing between posed and spontaneous smiles. The results showed that EMG has the advantage of being able to identify covert behavior not available through vision. Moreover, CV appears to be able to identify visible dynamic features that human judges cannot account for. This sheds light on the role of non-observable behavior in distinguishing affect-related smiles from polite positive affect displays.

CCS CONCEPTS

• **Human-centered computing** → **User models**;

KEYWORDS

Facial expression recognition; Computer vision; Electromyography.

ACM Reference Format:
Monica Perusquía-Hernández, Saho Ayabe-Kanamura, Kenji Suzuki, and Shiro Kumano. 2019. The Invisible Potential of Facial Electromyography: A Comparison of EMG and Computer Vision when Distinguishing Posed from Spontaneous Smiles. In *CHI Conference on Human Factors in Computing Systems Proceedings (CHI 2019), May 4–9, 2019, Glasgow, Scotland Uk.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3290605.3300379

CHI 2019, May 4–9, 2019, Glasgow, Scotland Uk

ACM ISBN 978-1-4503-5970-2/19/05…$15.00
https://doi.org/10.1145/3290605.3300379

1 INTRODUCTION

Measuring positive experiences in a continuous and accurate manner is of utmost importance. They are a metric of success in our personal life and in product and service design. In a user-centered design process, design concepts are drawn from existing user needs or bad experiences. The concept is then prototyped, and the new user experience is assessed with a view to improving it [24, 29]. Hence, unbiased user feedback is critical. This assessment is often performed via qualitative or quantitative self-report methods. To assess the dynamics of mental health and user experience it is essential to measure the frequency and patterning of mental processes in every-day-life situations, including affective experiences [10]. The Experience Sampling Method (ESM) [28] provides a good approximation of what the user is feeling when using a product. However, it requires a logging tool that intermittently prompts users to report their experience, and that ultimately alters the user's affective state.

An alternative way of assessing human affect in a continuous manner is to measure and interpret behavior and electrophysiological cues automatically using artificial intelligence (AI) technology. These methods have the potential to provide uninterrupted, objective measurements with high temporal resolution. The continuous logging of affective experience could trigger qualitative ESM entries or the reaction of an artificial agent after an affective experience has been identified. Perhaps the most widely used measurement modality is computer vision (CV), followed by electromyography (EMG) and other sensors for measuring autonomic body responses [6, 23, 47]. CV and EMG are used to identify and quantify behavior co-occurring with affective experiences such as facial expressions. However, humans are able to display positive facial expressions even when they are not experiencing any emotion. In particular, when evaluating a product or service, avoiding polite displays of positive experiences is desirable to better identify points of opportunity. Thus, it is also important to distinguish between posed and spontaneous facial expressions of positive affect to avoid bias in the evaluation process.

Positive affect is prototypically expressed in the form of a smile. According to the Facial Action Coding System (FACS) [14], a smile is often a combination of a lip corner puller (AU12) and a cheek raiser (AU06). AU06 corresponds to the activation of the *orbicularis oculi* muscle, and it is referred to as "the Duchenne Marker". Smiles with the Duchenne marker have been assumed to be spontaneous [13]. This definition has been used in numerous psychological studies [3, 17, 42]. However, it has also been shown that this muscle is activated in both types of smiles [32, 45]. Recently, more reliable differentiating features have been found. Dynamic aspects of facial expressions have been deemed critical for human perception. In particular, for subtle expressions, and when static information is of low quality [26]. Moreover, spontaneous smiles tend to last longer than posed smiles [8, 39], and have a fast and smooth onset [40], and apex coordination [15]. On the other hand, posed smiles have a larger amplitude [8, 39, 40]; different decay and rise durations and speeds [20, 30, 39], and different numbers of peaks [30].

The potential of computer vision (CV) [12] and electromyography (EMG) [37] to distinguish between these smiles has already been shown. However, little is known about how these two technologies relate to each other. Simultaneous recordings have so far been to the detriment of CV. Traditionally, EMG research requires electrodes to be placed on top of the relevant muscle. This not only prevents the wearer from producing natural facial expressions, it also obstructs recordings of the face. This limitation has recently been overcome by the use of wearables for measuring facial distal EMG [7, 18].

We present a direct comparison of these two methods to identify posed and spontaneous smiles. We hypothesize that they have different strengths. CV might be able to better discriminate overt facial expressions with fine spatial resolution, similar to human perception. On the other hand, EMG-based methods are potentially better at discriminating covert expression changes not perceivable through vision [44]. Specifically, we hypothesize an information space as depicted in Figure 1. This space is described by using four hypotheses: (1) EMG-based identification is superior for covert behavior; (2) CV-based identification is superior for overt behavior where spatial resolution is important; (3) the performance of a human judge is similar to that of CV, as it is based on visible behavior; (4) there is a shared information zone where all the proposed methods perform similarly.

2 RELATED WORK

Several surveys have summarized the methods available for automatic emotion recognition [6, 23, 47]. However, it is difficult to compare them due to the plethora of experimental paradigms, signal types, features, and classification schemes used to identify different combinations of emotions. Additionally, studies using simultaneous EMG and CV are rare due to the facial occlusion caused by traditional EMG.

Computer vision-based methods

Computer vision (CV) is the most widely used technique for identifying facial expressions or posture automatically [4]. The use of spatial patterns has been shown to achieve about 90% accuracy in the task of distinguishing between posed and spontaneous smiles [48]. In particular, the publication of the UvA-NEMO database including 1240 videos of spontaneous and posed smiles [12] has triggered a renewed interest in identifying the differences between posed and spontaneous smiles and their dynamic characteristics [19]. State-of-the-art methods have provided an identification accuracy up to 92.90% by using dynamic features based on lip and eye landmark movements, sometimes tailored to different age groups [12]. Other algorithms using spatio-temporal features as identified by restricted Boltzmann machines have been able to achieve up to 97.34% accuracy on the UvA-NEMO database, and 86.32% in the Spontaneous vs. Posed Facial Expression (SPOS) database [49].

Electromyography-based methods

The potential for using EMG to study different facial expressions has been widely reported. This is accomplished either by placing recording electrodes on top of the relevant muscle [5, 15, 33, 34, 41, 44, 46], or with wearable devices that do not obstruct the face [7, 18]. Posed and spontaneous smiles can also be distinguished by examining different EMG features. Surface EMG has revealed that spontaneous smiles have different magnitudes, speeds and durations [8, 40]. Furthermore, when using a wearable device with distal EMG [37], spatial and magnitude feature analysis allows us to distinguish between spontaneous and posed smiles with an accuracy of about 74%. By employing spatio-temporal features, the accuracy reached about 90%.

Comparison of CV and EMG

EMG has long been a promising technology for measuring unobservable facial behavior [44]. Simultaneous EMG and video recordings have shown that EMG onsets occur 0.23 s before lip corner movement. This makes EMG suitable for studying the fast reactions involved in posed and spontaneous expressions of emotion, which are believed to differ [8, 40, 41]. However, in these studies, the facial movements were somewhat restricted, and trials that included occlusion and head movements were discarded. These limitations can be overcome by using distal EMG, which does not obstruct the face. All in all, each technology might prove useful depending on the intended application. Although wearable EMG only exerts slight pressure on the sides of the face, CV is definitely less obtrusive than EMG because it does not require users to wear anything. However, CV is not robust to occlusion, head rotation, poor lighting or sudden movement.

Human Perception

Understanding another's facial expressions is a critical social ability. Therefore, human perception of facial expressions has been extensively researched. It has also been argued that the message transmitted by each facial expression is as important as the actual ground truth under which they were elicited [11], as they transmit both biologically basic and socially specific messages [22]. Perception has also been shown to be closely linked to facial mimicry. When distinguishing between posed and spontaneous smiles, the facial mimicry reactions of the perceiver were stronger for Duchenne smiles [25]. Additionally, it has been argued that the presentation of dynamic stimuli significantly enhances the human discrimination of posed and spontaneous smiles [26, 31]. Similar findings have been reported for facial expressions of surprise. The discrimination accuracy of human judges trying to distinguish between spontaneous and both improvised and rehearsed posed expressions was enhanced with dynamic stimuli. Nevertheless, only around 50% accuracy was achieved [50].

3 DATASET

Video clips and EMG data of posed and spontaneous smiles recorded in [35] were used for data analysis. In addition, the data they reported on the perception of these smiles by human judges was used as a reference. This data set was chosen because, to the best of our knowledge, there are no other data sets available that contain facial videos and facial distal EMG recorded simultaneously during displays of posed and spontaneous smiles.

Participants. 38 volunteers took part in the study (19 female, average age = 25.03 years old, SD = 3.83). Henceforth, these participants are referred to as "Producers" since they produced the smiles used for identification.

Experimental Design. The experiment consisted of three blocks. The first block or "Spontaneous Block" (S-B) was designed to induce a positive affective state, and therefore, elicit spontaneous smiles. S-B consisted of three 30 s silent humorous video clips presented in a counterbalanced order. The participants subsequently completed an Affect Grid [38] reporting how they felt while watching the videos. Then they video-coded their own facial expressions, indicating if the expression was posed or not. The second block or "Neutral Block" was designed to return the previously elicited positive affect to a neutral affect by having the participants watch a neutral-valenced video. The N-B video consisted of 18 pictures with likeability scores between 5.0 and 6.0 taken from the International Affective Picture System (IAPS) [27], presented every 5 s, for a total of 90 s. Next, in the third block or "Posed Block" (P-B), participants were asked to make similar facial expressions to those that they made when watching the first video in such a way that another person would not be able to guess whether or not the smiles were genuine. This type of smile was considered to be a posed smile that was intended to convey the impression of having fun. The P-B stimuli video consisted of 18 IAPS pictures with likeability scores between 4.0 and 5.0, presented every 5 s. In other words, the participants had to watch a slightly unpleasant stimulus while trying to fake smiles of enjoyment. Next, they video-coded their own facial expressions. All the participants experienced the experiment blocks in the same order.

Measurements. Videos of the participants' facial expressions were recorded using a Canon Ivis 52 at 30FPS. A wearable device with four unobtrusive channels of distal facial EMG was used (Figure 1). This device has been shown to reliably identify smiles in different situations [16, 18, 36, 43].

Collected data. According to the Affect Grid answers, valence scores were significantly higher in the spontaneous block than in the posed block ($F(1,64) = 11.47$, $p < .01$, $\eta_p^2 = 0.64$). This suggests that the producers felt more positive in the spontaneous block, and that they had to smile in the posed block even if they had slightly negative feelings. On the other hand, arousal did not differ among the experiment blocks ($F(1,64) = 0.50$, $p > .05$, $\eta_p^2 = 0.22$). According to their own video coding, 272 smiles were elicited from 32 participants. 127 were labeled as spontaneous and occurred in the S-B. 145 were posed and occurred in the P-B. In addition to the participant's own video coding, two independent raters labeled the videos. When judging whether participants were smiling or not, the Fleiss' Kappa indicating the agreement between the two coders and the participant's own video coding was 0.57. However, the agreement fell to to 0.13 when the task was to determine whether the displayed expressions were posed or spontaneous. Therefore, both experimental design and self-report were considered when establishing the ground truth labels. For a smile to be spontaneous it should be labeled as spontaneous by the participant and have occurred in the S-B. Similarly, for posed smiles, only those labeled as posed and occurring in the P-B were selected. Following this criterion, only 27 of the participants showed at least two smiles of each type.

In this dataset, the most relevant features are magnitude and rise and decay speed. These are calculated from a neutral expression to the apex of the smile and back to a neutral expression as measured from the EMG activity [35]. It is important to notice that the elicited posed smiles are different from those in previous studies. The most common elicitation paradigm is to directly ask the producers to smile. However, when posed smiles are requested with a command, the temporal dynamics are affected by the duration of the command itself. Dynamic information is critical for both human and machine perception [26, 50]. Hence, other more ecologically valid posed smiles were elicited. The producers voluntarily produced a smile that did not match the elicited affective inner state.

Human Judgment. 73 volunteers unknown to the producers took part in a separate study (37 female, average age = 29 years old, SD = 11). Henceforth, these participants are called "Perceivers". The perceivers watched 54 video clips of smiles (27 posed smiles and 27 spontaneous smiles) selected from smiles elicited as described above. All the perceivers watched all the selected smiles only once in random order. After watching each video clip, the perceivers labeled each smile as spontaneous or posed. A one sample t-test showed that their labeling accuracy was significantly different from chance level ($M = 0.50$, $t(72) = 8.80$, $p < .01$, 95% CI [.56 .59], $d = 0.58$).

Table 1: Features used for comparison.

	Spatial Features	**Spatio-temporal Features**
CV	-Intensity AU06 -Intensity AU12	-Duration -Rise and decay speed -Magnitude
EMG	-IC RMS from four EMG channels	-Duration -Rise and decay speed -Magnitude

4 ANALYSIS

The data obtained from the 27 producers who produced at least two smiles of each type was selected for analysis. The main aim of the analysis was to compare the EMG- and CV-based methods. Therefore, simple algorithms applicable to both methodologies were preferred over more complex classifiers.

Two types of features were used for each of the two modalities (Table 1). First, spatial features are magnitude metrics calculated sample by sample, independently of the number of smiles. Second, spatio-temporal features are obtained from each smile sample. Thus, the total number of data points available for the spatial algorithm (1,313,528 times four channels for EMG, and 230,342 times two AUs for CV for all smiles and all subjects) is higher than the data available for the spatio-temporal version (245 smiles of both types by four features by all subjects).

Feature extraction EMG. A similar algorithm to that described in [37] was used to calculate both the spatial and the spatio-temporal features of different smiles. In both cases the surface EMG was recorded from four channels at a sampling rate of 1 kHz . First, the data was band-pass filtered from 5 to 350 Hz. Second, it was notch filtered at harmonics of 50 Hz up to 350 Hz. Third, the EMG signal was linearly detrended to prevent drifts in the signal from contributing excessively to classifier performance. Next, the signals were decomposed using Independent Component Analysis (ICA) [9, 21] to separate the distal EMG from different source muscles.

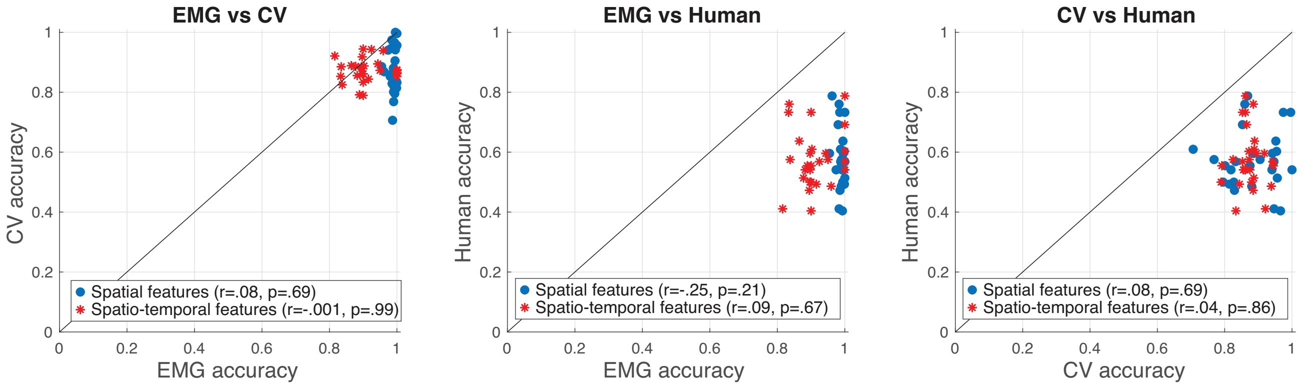

Figure 2: Classification results achieved with computer vision (CV) and electromyography (EMG) with spatial and spatio-temporal features, and by human judges. EMG seems able to identify covert behavior that is not visually discernible as suggested by its performance. CV seems able to identify visible dynamics that human judges could not account for. Asterisks represent significant differences in the average accuracies.

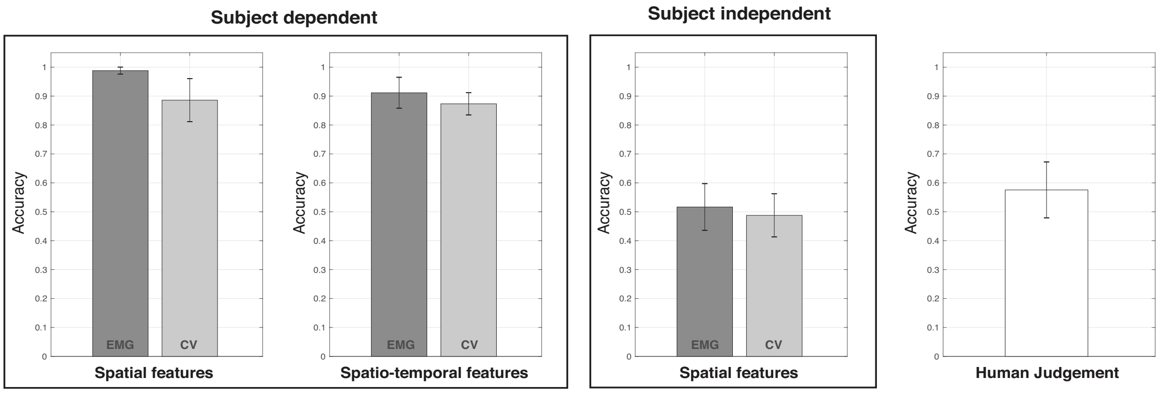

Figure 3: Average accuracy achieved by CV and EMG with spatial and spatio-temporal features, and by human judges. The spatial features model was also trained in a subject independent cross-validation. The performance is very similar across modalities, both in the subject-dependent and subject-independent models. The performance deteriorates for a subject independent model, which suggests marked individual differences probably due to differences in smiling style and muscle anatomy.

(1) Spatial features. (1) Absolute value was applied to each independent component (IC). Then, their root-mean square (RMS) was calculated using overlapping 100 ms windows and sliding one sample at a time.
(2) Spatio-temporal features. An envelope was fitted to the rectified ICs by smoothing the data with an averaging non-overlapping window of 100 ms, and a Savitzky-Golay filter with a 5th order polynomial and with 41 as the frame length. Then the maximum and minimum points of the envelope were identified. The maximum magnitude, rise time, decay time, rise speed, decay speed, and duration of the smile were calculated using such peaks.

Feature extraction CV. Basis features were calculated using OpenFace Toolkit 2.0 [1], a deep-learning, state-of-the-art facial analyzer.

(1) Spatial features. The intensities of AU6 and AU12 calculated frame by frame by OpenFace 2.0 were used as input for the classifier.
(2) Spatio-temporal features. Landmarks for lips and eye corners as calculated by OpenFace 2.0 were corrected for head orientation in each frame. Then the displacement of relevant landmarks with respect to the first frame of the video was calculated. This is the procedure for 3D landmark correction described in [19]. The lip landmarks we used were 48 and 54. We chose eye landmarks of 37, 40, 44, and 47. Landmark displacement was then smoothed in a similar fashion as with the EMG algorithm. Only the averaging window was

slid one frame at a time, to avoid data undersampling. Analogously, peak detection was performed, and the features of maximum magnitude, rise time, decay time, rise speed, decay speed, and duration of the smile were calculated and used as input for the classification.

Classification. Not all participants displayed the same number of posed and spontaneous smiles, and the smile durations varied greatly. Therefore, the feature vectors of the majority class were undersampled to match the size of the minority class. We used a support vector machine (SVM), with a radial basis kernel, trained with each feature set. The best hyperparameter set was chosen automatically with the fitcsvm Matlab function, independently for each modality. In both the EMG and CV models, subject dependent cross-validation was used with 85% training and 15% validation data. A test set was not separated from the cross-validation train and validation sets in the subject-dependent models because of the limited number of smiles of some of the producers. Additionally, subject independent SVM was fitted in a cross-validation and used with 85% training and 15% validation data from the N-1 producers. The data of one producer was left out for use in testing the SVM model built with the other N-1 producer's data. This procedure was repeated for all the producers. The reported subject-independent accuracy is the accuracy achieved when each producer was the test producer. Finally, Spearman's correlation coefficients between the results of each modality and feature type were calculated.

5 RESULTS

The mean accuracy of the subject-dependent spatial features was 88% (SD = 7%) for CV, and 99% (SD = 1%) for EMG. With the subject-dependent spatio-temporal features, it was 87% (SD = 4%) for CV, and 91% (SD = 5%) for EMG. The accuracy of the classification of each producer for each model is shown in Figure 2. As regards human judgment, the accuracy of each point is the average for all the judges (perceivers) per target (producers). From the plot it can be observed that, for the spatial features, EMG accuracy is high even when the accuracy with CV is lower. Wilcoxon signed-rank tests indicated that the difference between EMG and CV is significant for both the spatial ($V = 0$, $p < .01$) and the spatio-temporal features ($V = 83.5$, $p < .05$). Similar tests were used to assess the differences between feature-type per modality. With EMG, the spatial features performed significantly better than the spatio-temporal features ($V = 293.5$, $p < .01$). However, the difference was not significant for CV ($V = 0$, $p > .05$). Moreover, EMG accuracy is consistently higher than human accuracy, both in the magnitude feature space ($V = 378$, $p < .01$) and the spatio-temporal one ($V = 0$, $p < .01$). CV appears more consistent with human perception in both types of features. However, the differences in accuracy are also significant in both the spatial ($V = 378$, $p < .01$) and spatio-temporal ($V = 0$, $p < .01$) cases. None of the cross-modality pairs showed a strong correlation for either spatial or spatio-temporal features ($|r| < .26$, $p > .10$).

The mean accuracy of the subject-independent spatial features model was 52% (SD = 8%) for CV, and 49% (SD = 7%) for EMG in the test set (Figure 3).

6 DISCUSSION

This paper aimed to shed light on information available through both CV and EMG measurements, and to compare it with human perception of the same data. For this purpose, comparable features for both modalities were used. Additionally, the same SVM algorithm with subject-dependent cross-validation was applied for both modalities (Figure 2 and 3). EMG appears to be the best performing modality, probably because it picks up covert behaviors that are invisible to the naked eye. Muscle activation can occur that inhibits facial movement [44], and therefore no visible information is available. Moreover, the performance of the EMG classifiers achieved with this data set appears opposite to that achieved in earlier work [37]. Previously, spatio-temporal features outperformed spatial features. This might be because the posed smiles elicited in this experiment are smiles faking a positive valence even if their self-report described a slightly negative valence. These smiles might be different from smiles elicited under instruction from an experimenter or smiles produced voluntarily for the camera. The magnitudes of the posed smiles elicited here appear weaker. Additionally, when smiles are performed based on an instruction, the instruction itself might alter the duration and other temporal dynamics of the smile. The elicitation method used in this data set carefully avoided this problem.

With the spatio-temporal features, the data is limited to the number of smiles shown, which is independent of the modality. Therefore, CV and EMG perform very similarly. On the other hand, with the spatial features, the amount of data is dependent on the sampling rate of the sensor used. Whereas EMG was sampled at 1kHz, the video data contained only 30 FPS. EMG might have had an advantage by measuring more samples per smile, increasing the difference in performance between CV and EMG. Future work should assess whether or not the performance of EMG-based distinction is affected if the EMG data is undersampled to match the camera's speed.

Even though CV and human judges use the same visual features, CV outperforms human judges in distinguishing posed and spontaneous smiles. Moreover, the EMG performance is constantly higher than human performance for the subject-dependent model. The lack of correlation between the modality-feature pairs in the subject-dependent models suggests that indeed EMG and CV are complementary. The fact that there is no high correlation between CV and humans probably suggests that humans rely on context and movements other than AU06 and AU12. The smile video clips shown to the perceivers lack other contextual information on which humans usually rely to make their judgments. Finally, the ground-truth labeling was based on tags assigned by the producers. Since they are not expert coders, they might have mainly tagged expression apexes. Therefore, the smile video clips might have lacked certain critical dynamics related to the start and end of the smile for perceivers to achieve a similar result to CV. On the other hand, the subject-dependent automatic models were tailored to each producer. Human judges did not know the producers, and they received no feedback during the task. Thus, the model they used to make their decisions depended on their previous experiences with other people. This might explain their modest performance compared with automatic recognition. This is also in line with the results obtained with subject independent models. The individual differences in smiling style and muscle anatomy are possible reasons for the drop in accuracy to roughly chance level. EMG is known to be dependent on each person's physiognomy, therefore large individual differences are expected. However, the low CV performance is surprising, given that other studies have achieved very high performances independently of the person smiling [12, 49].

All in all, CV and EMG appear to be measuring complementary information that can be useful depending on the situation. Whereas CV is easy to set, non-obstructive, and sufficiently accurate, state-of-the-art algorithms are too computationally expensive to be useful in online settings or with embedded devices. On the other hand, traditional EMG is too obstructive to be of any practical use during user evaluations. However, wearable technologies capable of measuring EMG distally have improved EMG's usability. Now it is possible to quantify user experience with minimum obtrusion. Moreover, wearable EMG technology might be advantageous in situations where facial expressions are suppressed by the wearer, when there is a high degree of movement and occlusion, and when high performance is required in an online setting. As wearable technologies improve, EMG technologies may become increasingly relevant for assessing behaviors imperceptible to CV or humans.

Finally, this study suggested that, at least in the subject-dependent case, a simple classifier suffices. The use of simpler, faster algorithms will become more relevant as we move to real-time, embedded identification applications. For example, smarter experience sampling intervals for qualitative ESM diary entries could be triggered by automatically detected significant events.

7 LIMITATIONS OF THIS STUDY

This work presented only one method of comparing the EMG- and CV-based identification of posed and spontaneous smiles. Many other methods can be tried for both modalities with a view to improving performance [2]. However, we strove to make the comparison as fair as possible by using the same features and classification methods. It is difficult to achieve a good comparison from the literature due to methodological differences. This is the first study to directly compare CV and distal EMG. By using a similar classification technique for both modalities, we attempted to explore individual differences in smiling behavior and the visibility of the information selected by each modality.

Moreover, the EMG's high performance has to be carefully interpreted due to possible bias in the results. First, the test set was not separated from the cross-validation train and the validation sets in the subject-dependent models due to the limited amount of data. Thus, there is still a possibility that our model is overfitted. Second, the experimental design for gathering the data might have influenced the results, as the spontaneous block always preceded the posed block. This choice was made to ensure that the producer's smiles were spontaneous. To mitigate the effects of this block design, trends in the EMG signal were removed during the analysis. However, the order effect might have influenced the EMG models. Nevertheless, this possibility is small, as the results achieved by CV and EMG are similar. Furthermore, the EMG preprocessing is based on [18, 36, 37]. The reported accuracy is consistently around 90% for different types of smiles and in counterbalanced experimental designs.

8 CONCLUSIONS AND FUTURE WORK

We compared the performance of CV and EMG measurements in the task of distinguishing posed and spontaneous smiles. The highest performance was achieved with EMG and spatial features. This might be due to the nature of the posed smiles elicited in this data set. Hence, further research should explore differences between different types of posed smiles. Furthermore, the results showed that EMG has an advantage as regards identifying covert behavior not available through vision. Future work should explore in more detail the influence of factors such as sampling rate and training scheme. Moreover, CV appears able to identify visible dynamic features that human judges cannot account for. This sheds light on the role of non-observable behavior in distinguishing affect-related smiles from posed smiles to avoid bias during automatic user experience assessments.

REFERENCES

[1] Tadas Baltrusaitis, Amir Zadeh, Yao Chong Lim, and Louis-Philippe Morency. 2018. OpenFace 2.0: facial behavior analysis toolkit. In *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*. IEEE, 59–66.

[2] Marian Stewart Bartlett, Gwen C. Littlewort, Mark G. Frank, and Kang Lee. 2014. Automatic decoding of facial movements reveals deceptive pain expressions. *Current Biology* 24, 7 (mar 2014), 738–743. https://doi.org/10.1016/J.CUB.2014.02.009

[3] Michael J. Bernstein, Donald F. Sacco, Christina M. Brown, Steven G. Young, and Heather M. Claypool. 2010. A preference for genuine smiles following social exclusion. *Journal of Experimental Social Psychology* 46, 1 (2010), 196–199.

[4] Vinay Bettadapura. 2012. Face expression recognition and analysis: the state of the art. *CoRR* (2012), 1–27. arXiv:1203.6722

[5] John T. Cacioppo and Louis G. Tassinary. 1990. Inferring psychological significance from physiological signals. *American Psychologist* 45, 1 (1990), 16–28.

[6] Rafael A. Calvo and Sidney D. Mello. 2010. Affect detection: an interdisciplinary review of models, methods, and their applications. *IEEE Transactions on Affective Computing* 1, September (2010), 18–37.

[7] Yumiao Chen, Zhongliang Yang, and Jiangping Wang. 2015. Eyebrow emotional expression recognition using surface EMG signals. *Neurocomputing* 168 (2015), 871–879.

[8] J. F. Cohn and K.L. Schmidt. 2004. The timing of facial motion in posed and spontaneous smiles. *International Journal of Wavelets, Multiresolution and Information Processing* 2 (2004), 121–132.

[9] Pierre Comon. 1994. Independent component analysis, a new concept? *Signal Processing* 36, 36 (1994). http://mlsp.cs.cmu.edu/courses/fall2012/lectures/ICA.pdf

[10] Mihaly Csikszentmihalyi and Reed Larson. 2014. Validity and reliability of the Experience-Sampling Method. In *Flow and the Foundations of Positive Psychology*. Springer Netherlands, Dordrecht, 35–54.

[11] Amy Dawel, Luke Wright, Jessica Irons, Rachael Dumbleton, Romina Palermo, Richard O'Kearney, and Elinor McKone. 2017. Perceived emotion genuineness: normative ratings for popular facial expression stimuli and the development of perceived-as-genuine and perceived-as-fake sets. *Behavior Research Methods* 49, 4 (2017), 1539–1562.

[12] Hamdi Dibeklioglu, Albert Ali Salah, and Theo Gevers. 2015. Recognition of genuine smiles. *IEEE Transactions on Multimedia* 17, 3 (2015), 279–294.

[13] Guillaume-Benjamin Duchenne. 1862. *Mécanisme de la Physionomie Humaine*. Jules Renouard, Paris.

[14] Paul Ekman, Wallace Friesen, and Joseph Hager. 2002. FACS investigator's guide.

[15] Paul Ekman and Erika Rosenberg. 2005. *What the face reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS)* (second edition ed.). Oxford University Press. 1–20, 453–486 pages.

[16] Atsushi Funahashi, Anna Gruebler, Takeshi Aoki, Hideki Kadone, and Kenji Suzuki. 2014. 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* 44, 3 (2014), 685–693.

[17] Reuma Gadassi and Nilly Mor. 2016. Confusing acceptance and mere politeness: Depression and sensitivity to Duchenne smiles. *Journal of Behavior Therapy and Experimental Psychiatry* 50 (2016), 8–14.

[18] Anna Gruebler and Kenji Suzuki. 2014. Design of a Wearable Device for Reading Positive Expressions from Facial EMG Signals. *IEEE Transactions on Affective Computing* PP, 99 (2014), 1–1.

[19] Hui Guo, Xiao-hui Zhang, Jun Liang, and Wen-jing Yan. 2018. The dynamic features of lip corners in genuine and posed smiles. *Frontiers in psychology* 9, February (2018), 1–11.

[20] Mohammed Hoque, Louis Philippe Morency, and Rosalind W. Picard. 2011. Are you friendly or just polite? - Analysis of smiles in spontaneous face-to-face interactions. In *Affective Computing and Intelligent Interaction. Lecture Notes in Computer Science*, Sidney D'Mello (Ed.). Vol. 6974. Springer Berlin Heidelberg, 135–144.

[21] Aapo Hyvärinen and Erkki Oja. 2000. Independent component analysis: algorithms and applications. *Neural networks: the official journal of the International Neural Network Society* 13, 4-5 (2000), 411–30. https://doi.org/10.1016/S0893-6080(00)00026-5

[22] Rachael E. Jack, Oliver G.B. Garrod, and Philippe G. Schyns. 2014. Dynamic facial expressions of emotion transmit an evolving hierarchy of signals over time. *Current Biology* 24, 2 (2014), 187–192.

[23] Joris H. Janssen, Paul Tacken, J.J.G. (Gert-Jan) de Vries, Egon L. van den Broek, Joyce H.D.M. Westerink, Pim Haselager, and Wijnand A. IJsselsteijn. 2013. Machines outperform laypersons in recognizing emotions elicited by autobiographical recollection. *Human–Computer Interaction* 28, 6 (2013), 479–517. https://doi.org/10.1080/07370024.2012.755421

[24] Jussi P.P. Jokinen. 2015. Emotional user experience: traits, events, and states. *International Journal of Human Computer Studies* 76 (2015), 67–77.

[25] Eva G. Krumhuber, Katja U. Likowski, and Peter Weyers. 2014. Facial mimicry of spontaneous and deliberate Duchenne and Non-Duchenne smiles. *Journal of Nonverbal Behavior* 38, 1 (2014), 1–11.

[26] Eva G Krumhuber and Antony S. R. Manstead. 2013. Effects of dynamic aspects of facial expressions: a review. *Emotion Review* 5, 1 (2013), 41–46.

[27] P.J. Lang, M.M. Bradley, and B.N. Cuthbert. 2008. *International Affective Picture System (IAPS)*. Technical Report. University of Florida, Gainesville, FL. arXiv:0005-7916(93)E0016-Z

[28] Reed Larson and Mihaly Csikszentmihalyi. 1983. The Experience Sampling Method. *New Directions for Methodology of Social & Behavioral Science* 15 (1983), 41–56.

[29] Ji-Ye Mao, Karel Vredenburg, Paul W. Smith, and Tom Carey. 2005. The state of user-centered design practice. *Commun. ACM* 48, 3 (2005), 105–109.

[30] Mohammad Mavadati, Peyten Sanger, Mohammad H Mahoor, and S York Street. 2016. Extended DISFA dataset: investigating posed and spontaneous facial expressions. , 8 pages.

[31] Shushi Namba, Russell S. Kabir, Makoto Miyatani, and Takashi Nakao. 2018. Dynamic displays enhance the ability to discriminate genuine and posed facial expressions of emotion. *Frontiers in Psychology* 9 (2018), 672.

[32] Shushi Namba, Shoko Makihara, Russell S. Kabir, Makoto Miyatani, and Takashi Nakao. 2016. Spontaneous facial expressions are different from posed facial expressions: morphological properties and dynamic sequences. , 13 pages.

[33] Lindsay M. Oberman, Piotr Winkielman, and Vilayanur S. Ramachandran. 2007. Face to face: blocking facial mimicry can selectively impair recognition of emotional expressions. *Social neuroscience* 2, 3-4 (2007), 167–78.

[34] Lindsay M. Oberman, Piotr Winkielman, and Vilayanur S. Ramachandran. 2009. Slow echo: facial EMG evidence for the delay of spontaneous, but not voluntary, emotional mimicry in children with autism spectrum disorders. 4 (2009), 510–520. https://doi.org/10.1111/j.1467-7687.2008.00796.x

[35] Monica Perusquía-Hernández, Saho Ayabe-Kanamura, and Kenji Suzuki. 2018. Human perception and biosignal-based identification of posed and spontaneous smiles. *Manuscript in preparation.* (2018).

[36] Monica Perusquía-Hernández, Masakazu Hirokawa, and Kenji Suzuki. 2017. A wearable device for fast and subtle spontaneous smile recognition. *IEEE Transactions on Affective Computing* 8, 4 (2017), 522–533.

[37] Monica Perusquía-Hernández, Masakazu Hirokawa, and Kenji Suzuki. 2017. Spontaneous and posed smile recognition based on spatial and temporal patterns of facial EMG. In *Affective Computing and Intelligent Interaction*. 537–541.

[38] James A. Russell, Anna Weiss, and Gerald A. Mendelsohn. 1989. Affect Grid: a single-item scale of pleasure and arousal. *Journal of Personality and Social Psychology* 57, 3 (1989), 493–502.

[39] Karen Schmidt, Sharika Bhattacharya, and Rachel Denlinger. 2009. Comparison of deliberate and spontaneous facial movement in smiles and eyebrow raises. *Nonverbal Behaviour* 33, 1 (2009), 35–45.

[40] Karen L. Schmidt, Zara Ambadar, Jeffrey F. Cohn, and L. Ian Reed. 2006. Movement differences between deliberate and spontaneous facial expressions: zygomaticus major action in smiling. *Journal of Nonverbal Behavior* 30, 1 (2006), 37–52.

[41] K. L. Schmidt and J. F. Cohn. 2001. Dynamics of facial expression: normative characteristics and individual differences. In *IEEE Proceedings of International Conference on Multimedia and Expo*. IEEE, Tokyo, 728–731.

[42] Ruiting Song, Harriet Over, and Malinda Carpenter. 2016. Young children discriminate genuine from fake smiles and expect people displaying genuine smiles to be more prosocial. *Evolution and Human Behavior* 37, 6 (2016), 490–501.

[43] Yuji Takano and Kenji Suzuki. 2014. Affective communication aid using wearable devices based on biosignals. In *Proceedings of the 2014 conference on Interaction design and children - IDC '14*. ACM Press, New York, New York, USA, 213–216.

[44] Louis G. Tassinary and John T. Cacioppo. 1992. Unobservable Facial Actions and Emotion. *Psychological Science* 3, 1 (1992), 28–33.

[45] Pascal Thibault, Manon Levesque, Pierre Gosselin, and Ursula Hess. 2012. The Duchenne marker is not a universal signal of smile authenticity - but it can be learned! *Social Psychology* 43, 4 (2012), 215–221. arXiv:arXiv:1011.1669v3

[46] Anton van Boxtel. 2010. Facial EMG as a tool for inferring affective states. In *Proceedings of Measuring Behavior*, AJ Spink, F Grieco, Krips OE, LWS Loijens, LPJJ Noldus, and PH Zimmerman (Eds.). Eindhoven, 104–108.

[47] Alessandro Vinciarelli, Maja Pantic, and Herve Bourlard. 2009. Social signal processing: survey of an emerging domain. *Image and Vision Computing* 27, 12 (2009), 1743–1759.

[48] Shangfei Wang, Chongliang Wu, and Qiang Ji. 2016. Capturing global spatial patterns for distinguishing posed and spontaneous expressions. *Computer Vision and Image Understanding* 147 (jun 2016), 69–76. https://doi.org/10.1016/J.CVIU.2015.08.007

[49] Jiajia Yang and Shangfei Wang. 2017. Capturing spatial and temporal patterns for distinguishing between posed and spontaneous expressions. In *Proceedings of the 2017 ACM on Multimedia Conference - MM '17*. ACM Press, New York, New York, USA, 469–477.

[50] Mircea Zloteanu, Eva G. Krumhuber, and Daniel C. Richardson. 2018. Detecting genuine and deliberate displays of surprise in static and dynamic faces. *Frontiers in Psychology* 9 (2018), 1184.