Facial movement synergies and Action Unit detection from distal wearable Electromyography and Computer Vision

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

Felix Dollack
NTT Communication Science Laboratories
University of Tsukuba

Chun Kwang Tan
University of Tsukuba

Saho Ayabe-Kanamura
University of Tsukuba

Kenji Suzuki
University of Tsukuba

ABSTRACT

Distal facial Electromyography (EMG) can be used to detect smiles and frowns with reasonable accuracy, by capitalizing on volume conduction to detect relevant muscle activity, even when the electrodes are not placed directly on the source muscle. The main advantage of this method is to prevent occlusion and obstruction of the facial expression production, whilst allowing EMG measurements. However, measuring EMG distally entails that the exact source of the facial movement is unknown. We propose a novel method to estimate specific Facial Action Units (AUs) from distal facial EMG and Computer Vision (CV). This method is based on Independent Component Analysis (ICA), Non-Negative Matrix Factorization (NNMF), and sorting of the resulting components to determine which is the most likely to correspond to each CV-labeled action unit (AU). Performance on the detection of AU06 (Orbicularis Oculi, ^_^) and AU12 (Zygomaticus Major, :)) was estimated by calculating the agreement with Human Coders. The results showed an accuracy of 83%, and a Cohen's Kappa of 0.42 for AU6 and 0.43 for AU12. This demonstrates the potential of distal EMG to detect individual facial movements. Using this multimodal method, several AU synergies were identified. Finally, we also quantified the co-occurrence and timing of AU6 and AU12 in posed and spontaneous smiles.

CCS CONCEPTS

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

KEYWORDS

electromyography, computer vision, smiles, FACS

ACM Reference Format:
Monica Perusquía-Hernández, Felix Dollack, Chun Kwang Tan, Saho Ayabe-Kanamura, and Kenji Suzuki. 2018. Facial movement synergies and Action Unit detection from distal wearable Electromyography and Computer Vision. In *Woodstock '18: ACM Symposium on Neural Gaze Detection, June 03–05, 2018, Woodstock, NY*. ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/1122445.1122456

ACM ISBN 978-1-4503-XXXX-X/18/06…$15.00
https://doi.org/10.1145/1122445.1122456

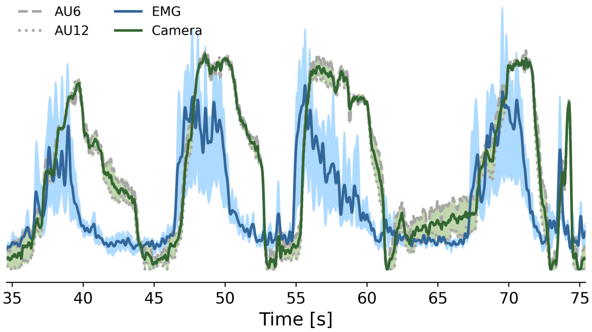

Figure 1: Muscle activations measured with distal EMG precede camera-detected AU movements. The four channels of raw distal EMG activate on average 374 ms before the detected CV-based AU labels. The plot shows data from one participant posing smiles. The activation patterns of EMG and CV-based AUs are similar to each other, with EMG activity leading. Finally, AU6 and AU12 often co-occur.

1 INTRODUCTION

Facial expressions often co-occur with affective experiences. Smiles are among the most ubiquitous facial expressions. They are characterized by the corner of the lips moving upwards, as a resulting action of the activity of the *Zygomaticus Major* muscle (ZM). Although smiles have been deemed the prototypical expression of happiness, not all people who smile are happy. The so-called Duchenne marker, or movement from the *Orbicularis Oculi* muscle (OO), often co-occurs with the *Zygomaticus Major* activity. Whilst it has been claimed that the Duchenne marker is a signal of smile spontaneity [8, 9, 12, 16], other studies have found this marker in posed smiles as well [23, 27, 33]. Moreover, others suggested that it might also signal smile intensity instead of smile authenticity [14, 23, 26].

The Facial Action Coding System (FACS) [11] is a standardized method to label facial movements. It involves identifying Action Descriptors for movements involving multiple muscles from their onset to offset. The advantage of using the FACS is that no subjective inferences about the underlying emotion are made during the facial movement identification. From these AU configurations, inferences can be made by experts in the frame of different theories of emotion. In the FACS, the lip corner pulling upwards is labeled as Action Unit 12 (AU12), and the movement around the eyes in the form of a cheek raiser is labeled as AU6. AU6 is also the AU associated with the Duchenne Marker. AUs can be measured by visual inspection using video recordings, either by a human coder [10] or by using Computer Vision (CV) algorithms [1]. Additionally, the underlying muscle activity can be measured with Electromyography (EMG) [34, 41]. The standard method is to place the EMG electrodes directly on top of the relevant muscle to increase Signal-to-Noise Ratio (SNR). More recently, several studies have proved the feasibility of measuring facial expressions with distal EMG [15, 17]. Distal EMG refers to measuring muscle activity from a body location that is distant from the relevant muscle. Distal EMG measurements are possible through volume conduction whereby the electrical activity generated by each muscle spreads to adjacent areas [41]. By measuring EMG distally, the unnatural obstruction that the electrodes pose to the production of facial expressions is reduced. Despite this advantage, distal measurements make it difficult to know the exact location of the EMG source. Hence, current technology has been only used to identify grouped muscle activity such as smiles or frowns. Detecting such facial expressions from EMG has its own merit, such as high temporal resolution and robustness against occlusion. However, only if movement activity is identified at the AU level, we would be able to compare the knowledge drawn using this technology to the large body of facial expression research that uses AUs as the basis of analysis.

Moreover, units of movement are often grouped to form full facial expressions. In movement science, muscle synergies are defined as joint movements produced by muscle groups [40]. Analogously, we define facial movement synergies as groups of muscles, or AUs, moving together. We hypothesize that if AU6 and AU12 move together in spontaneous smiles, but not in posed smiles, we should be able to observe different synergies as described by either muscle activity, CV-labeled AUs or human-labeled AUs. On the other hand, if the Duchenne marker is not a signature of spontaneity, we should observe similar synergies in both posed and spontaneous smiles. For the purpose of this analysis, we propose a sensing-source-synergy framework. Blind Source Sepparation (BSS) methods, such as Independent Component Analysis (ICA) or pre-trained OpenFace models, could be used to go from sensing to sources. Furthermore, analyses such as Non-Negative Matrix Factorization (NNMF) can be used to identify synergies from fine-grained movement units. We explore several algorithms within this space to shed light on the spatial and temporal elements that form posed and spontaneous smiles.

The main contributions of this work are: **(1) A framework** to analyze sensed signals by estimating their sources and synergies. **(2) a method to identify individual muscle activity sources** linked to the AUs 6 and 12 during smile production from a multimodal system. This system uses both CV and EMG during calibration, and EMG only for high-movement, high-occlusion situations. We use CV as automatic labeling method to identify different EMG components. **(3) An analysis of facial movement synergies** using AUs as sources. We present two selection methods to analyze activation patterns of AU6 and AU12 in the context of posed and spontaneous smiles.

2 RELATED WORK

2.1 EMG-based identification

Compared to traditional EMG measurements, a reduced set of electrode positions has proven to yield high facial expression recognition rates of 87% accuracy for seven posed facial expressions, including sadness, anger, disgust, fear, happiness, surprise and neutral expressions. This subset includes electrodes placed on the *Corrugator* and *Frontalis* on the forehead; and ZM and *Masseter* on the cheeks [35]. Distal EMG has been used to identify different facial gestures by using different electrode configurations. Two EMG bipolar channels were placed on the *Temporalis* muscle on each side of the face, and one placed on the *Frontalis* muscle gave input to distinguish ten facial expressions. The achieved accuracy was 87% using a very fast versatile elliptic basis function neural network (VEBFNN) [17]. Although not all gestures were facial expressions of emotion, they did include symmetrical and asymmetrical smiling, raising eyebrows, and frowning. Moreover, distal EMG has been implemented on a wearable designed to keep four EMG channels attached to the sides of the face at eye level. With this placement, it is possible to reliably measure smiles in different situations without obstructing facial movement [13, 31]. This is possible because smile-related distal activity measured from the ZM is sufficiently large to be robust against non-affective facial movements such as chewing gum and biting [15, 28, 41]. Hence, the information picked up by the four channels is used to approximate different sources of muscular activity using Independent Component Analysis (ICA) [6]. The separated muscle activity contains components for muscles involved in generating smiles and can be used to identify these [15]. This approach can be used offline for fast and subtle spontaneous smile identification [31] and is possible even in real time [36]. Finally, this device has also been used to analyze spatio-temporal features of a smile by fitting envelopes to the EMG's Independent Components (ICs), and later performing automatic peak detection on those envelopes [32] with performance similar to that achieved by Computer Vision [30]. Furthermore, four EMG leads placed around the eyes in a Head-Mounted Display (HMD) have also been used successfully to distally identify facial expressions even when the face is covered by the device. Facial expressions of anger, happiness, fear, sadness, surprise, neutral, clenching, kissing, asymmetric smiles, and frowning were identified with 85% of accuracy [3]. Another recent work proposed the use of a thin sticker-like hemifacial 16 electrode array to paste on one side of the face and identify ten distinct facial building blocks (FBB) of different voluntary smiles. Their electrode approach is novel, robust against occlusion, and provides a higher density electrode array than that of the aforementioned arrangements. This enabled them to use ICA and clustering to define several FBB corresponding to a certain muscle [21]. Nevertheless, they require electrode usage proximal to each muscle. This entails that a large sticker needs to be placed on the skin, obstructing spontaneous facial movement. Moreover, the physical connection of the electrode array enhances artifact cross-talk between electrodes. To eliminate such cross-talk, ICA was used and the resulting clusters were derived manually.

2.2 CV-based identification

CV is the most widely used technique for identifying facial expressions [2], even at the individual Action Units level. There are different approaches to extract relevant features for AU identification and intensity estimation. Among these, appearance-based, geometry-based, motion-based, and hybrid approaches. Several algorithms range between 0.45 and 0.57 F1 scores for occurrence detection and between 0.21 and 0.41 for intensity estimation [25]. The OpenFace toolkit 2.0 [1] is a CV pipeline for facial and head behavior identification. Its behavior analysis pipeline includes landmark detection, head pose and eye gaze estimation, and facial action unit recognition. This algorithm detects AU 1, 2, 4, 5, 6, 9, 12, 15, 17, 20, 25, 26 with an average accuracy of 0.59 in a person-independent model. Moreover, the use of spatial patterns has been shown to achieve about 90% accuracy in the task of distinguishing between posed and spontaneous smiles [42]. Dynamic features based on lip and eye landmark movements have provided an identification accuracy up to 92.90% [7]. Other algorithms using spatio-temporal features as identified by restricted Boltzmann machines have been able to achieve up to 97.34% accuracy in distinguishing spontaneous vs. posed facial expressions [43].

2.3 Muscle synergies

The concept of synergy in facial activity recognition has been used mainly to describe synergies between different sensors [18, 19, 22]. However, muscle synergies refer to simultaneous muscle activation. Muscle synergies are typically expressed in the form of a spatial component (synergies) and a temporal component (activations). The spatial component describes the the grouping and ratio of muscles that are activating together for a given movement. The temporal component describes how each spatial component is activated in the time series. Two types of synergies have been currently proposed: "Synchronous synergies" assume there are no temporal delays between the different muscles forming the synergies (i.e. synergies are consistent throughout the movement) while the activation of these components change. On the other hand, "time-varying synergies" consider both the synergies to change [40]. There is an ongoing debate on whether the Central Nervous System controls the activation of individual motor units, individual muscles, group of muscles, or kinematic and automatic features [40]. This research is mainly done in the domain of motor control of wide and coordinated movements such as gait. Several researchers have used NNMF as a method to identify muscle synergies during posture and gait responses. This method is often used jointly with an analysis of Variance-Accounted-For (VAF) by each synergy component when reconstructing the original EMG signal [37–39]. In this method, the source signals are decomposed in as many components as there are degrees of freedom. The number of components that contain a VAF higher than a threshold are considered as the number of synergies contained in the group of sources.

3 THE SENSING-SOURCE-SYNERGY FRAMEWORK

We propose a framework to analyze sensed signals by estimating their sources and synergies. Since AUs are closely related to individual muscle activity, we refer to them as "sources" (Fig. 2). Sources are facial movement units caused by a certain muscle. These individual sources often move in synchronous manner to form visible facial expressions such as smiles. A group of muscles, or a group of AUs, moving together are called synergies. Different transformations are necessary to go between sensed signals, source signals, and synergies. To go between a sensed signal mix to the source signals originating a movement, we can use ICA for EMG, and OpenFace for videos. Similarly, NNMF can be used to go from movement sources to synergy groups. A special case is when the sensed signal is very close to the source signal. For example, if EMG is measured directly from the muscle originating the muscle, we can assume equivalence between measurement and source. In such cases, NNMF can be applied directly to the sensed signal to obtain synergies.

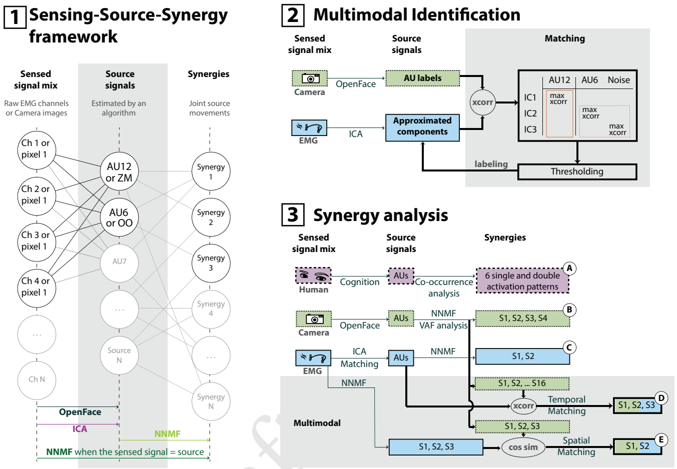

Figure 2: (1) Sensing-Source-Synergy framework. Sensors often read raw signals that are a mixture of signals of interest and other sources that can be considered noise. In other cases, the measured signals can be considered directly as the sources. We use OpenFace and ICA to derive movement source units. (2) Multimodal AU identification. AU labels are extracted from CV, and are used to assess facial expressions. Then the information derived from CV is used to identify the EMG components that correspond to each AU type. (3) Synergy analysis. Groups of sources moving together were analyzed by looking at the different sensing modalities independently or in combinations. (A) Six single and double activation patterns were proposed and identified from discrete AU labels. (B) Using NNMF and VAF methods, four AU synergies were found. (C) ICA was used to find the muscle sources and then NNMF was applied on the ICs to identify two synergies. (D) Temporal matching of each of the possible synergies with the ICs identified from EMG. (E) Spatial matching of the NNMF weights derived from EMG and four CV-labeled AUs selected from the VAF analysis.

4 DATA SET

This data is a subset of the data generated in a previous study [29]. Here only a brief description is provided for informative purposes.

4.1 Participants

41 producers took part in the study (19 female, average age=25.03 years, SD=3.83).

4.2 Experiment design

The experiment consisted of several blocks. All the producers completed all the experimental blocks in the same order. This was to keep the purpose of the experiment hidden during the spontaneous block.

4.2.1 Spontaneous Block (S-B). A positive affective state was induced using a 90 s humorous video. After the stimuli, a standardized scale assessing emotional experience was answered. Next, producers were asked to tag any facial expressions that they had made.

4.2.2 Posed Block (P-B). Producers were requested to make similar smiles as they did in the S-B. However, this time, a 90 s slightly negative video was presented instead. Their instruction was: “Please perform the smiles you video coded. This is for a contest. We are going to show the video we record to another person, who is unknown to you, and if she or he cannot guess what video you were watching, then you are a good actor. Please do your best to beat the evaluator". After watching the video and performing the task, they completed the same standardized scale assessing emotional experience. They were also asked to tag their own expressions.

4.3 Measurements

- **Smile-reader.** Four channels total of distal facial EMG were measured from both sides of the face using dry-active electrodes (Biolog DL4000, S&ME Inc) sampled at 1 kHz.
- **Video recordings.** A video of the producer’s facial expressions was recorded using a Canon Ivis 52 camera at 30 FPS.
- **Self video coding.** The producers tagged the onset and offset of their own facial expressions using Dartfish 3.2. They labeled each expression as spontaneous or posed, and indicated whether or not it was a smile.
- **Third person video coding.** Two independent raters labeled the videos with Dartfish 3.2. They coded for the start frame and the duration of every smile, and AUs 1, 2, 4, 5, 6, 9, 10, 12, 14, 15, 17, 18, 25, 26 and 28.

5 DATA ANALYSIS

Two types of data analysis were performed (Fig. 2). The first type is multimodal identification, which aims to identify different sources or muscle groups from the recorded EMG; and assesses their similitude to AUs detected using CV. The second one is synergy analysis, aimed to identify the spatial and temporal structures of the facial expressions present in the data. The synergy analyses were conducted both on single modalities and at the multimodal level. Both types of analysis used the same type of EMG pre-processing.

EMG Pre-processing. The four EMG channels were first passed through a custom Hanning window with a ramp time of 0.5 s to avoid introduction of artificial frequencies by the filtering at the start and the end of the signal. Afterwards, the signals were (1) linear detrended, (2) transformed to have zero mean and one standard deviation, (3) band-pass filtered from 15 to 490 Hz, (4) rectified, and (5) low-pass filtered at 4 Hz.

CV-based AU labeling using OpenFace. The Facial Behavior Analysis Toolkit OpenFace 2.0 was used to identify several facial features including AUs. AU identification is given both as a continuous output or intensity rating; and a binary output indicating AU presence. The intensity and presence predictors have been trained separately and on slightly different datasets, which means that they are not always consistent [1]. In this work, we choose to use the continuous or the binary rating depending on the requirements of our algorithm. The binary CV AU labels are extracted and upsampled, as well as the continuous labels, from 30 Hz to 1 kHz to match the EMG sampling frequency.

Blind-source separation. ICA [20], was used to automatically estimate different muscle activity sources. The wearable used to collect the data has four channels. Thus, we set the number of decomposed components to three.

Synergy identification. NNMF [24] is a dimensional reduction method aimed to uncover synchronized muscle movements. We expect it to be able to identify source activity happening at the same time, either from CV-AUs or EMG. If AU6 and AU12 happen at different times, they should be categorized as belonging to different synergies. The used dataset contains both posed and spontaneous smiles. Thus, we hypothesize that NNMF will be able to identify whether AU6 and AU12 belong to the same synergy or not. The number of synergies found primarily depends on the number of available sources. In the case of CV, 17 AUs were identified. Hence, the degrees of freedom were a maximum of 16. In the case of EMG, the degrees of freedom is three when applying NNMF on the raw measurements, and two when applied on the estimated ICs. Fig. 2-3 shows in detail the processing followed.

Multimodal identification with component matching to CV-generated labels. A matching method was used to assess similarity between EMG components and CV-based labels (Fig. 2-2). We assume the EMG signal to contain AU6, AU12 and noise. Noise is defined as electrical interference as well as other muscular sources. First, we calculated the cross-correlation of the three ICA components; the continuous AU6, AU12 OpenFace CV-labels; and an uniformly distributed random noise distribution. Since AU12 stems from the large and strong ZM muscle, the index of the maximum correlation is chosen to correspond to AU12. The other two ICs get assigned to be AU6 and noise in order of maximum correlation value. Afterwards, a threshold method was used to determine active samples. Further smoothing is applied on the individual ICs by means of a first order Savitzky-Golay filter with length 301. An initial period of $\approx 1\ s$ or 30 samples of the IC is used to calculate the baseline signal average and standard deviation. The whole signal then is turned into a binary vector where samples that cross the threshold of $\overline{m} + k\sigma$ with $k = 2$ are set to one. The values set to one are thought to correspond to activity of the respecting AU assigned to the IC during the process of AU identification.

Synergy Analysis. In some contexts, AUs might appear together more often. This might be the case for the co-occurrence of AU6 and AU12 in spontaneous smiles, should the Duchenne marker truly be a marker of smile spontaneity. In contrast, posed smiles would be characterized by less synchronized AU6 and AU12 activity. Hence, several co-occurring activation were analyzed using different modalities (Fig. 2-3-A). We propose six activation patterns: (1) AU6 only; (2) AU12 only; (3) AU12 inside AU6; (4) AU6 inside

AU12; (5) AU12 before AU6; and (6) AU6 before AU12. An algorithm was designed to detect these from binary labels. These labels can be generated by human coding, CV, EMG or a combination of them. This method relies on subtracting the labels of one AU from the other, and then identifying the differences within each block for each activation pattern.

Moreover, NNMF was used to extract muscle synergies from EMG, and AU synergies from the continuous CV-labels. The total variance accounted for (VAF) was used as a metric to determine the ideal number of synergies [37]. First, we used the AUs derived from CV (Fig. 2-3-B), and decomposed them into NNMF components between 1 and 16 per experimental block. Afterwards, the ideal number of synergies was determined by ensuring that the number of synergies would reconstruct the original signal with less than 15% error for all participants in both posed and spontaneous blocks . Moreover, we compared the EMG-based and CV-based synergy detection to match them across modalities. The matching was done in the temporal domain using cross-correlation (Fig. 2-3-D), and in the spatial domain using cosine similarity to sort and match the cross-modal NNMF component weights (Fig. 2-3-E). To use cosine similarity, it is necessary to have components with equal number of weights in both modalities. Thus, we decomposed three synergies from only four smile-related CV-labeled AUs (AU 6, 7, 10 and 12). This was to match the four EMG channels. Nevertheless, we would prefer to do a cascading transformation by first determining the sources using ICA, and then applying NNMF to the ICs (Fig. 2-3-C). However, due to the limited number of EMG channels and the evidence from the multimodal identification that AU12 and AU6 are strong in the raw signal; we considered the EMG signal to be equivalent to the sources only for this analysis (Fig. 2-3-E). Given this, the spatial matching would be an alternative to the multimodal identification described in Fig. 2-2.

Delay between raw EMG and CV-based AU detection. The delay between the EMG signals and the CV-based AU detection was calculated by looking at the lag at which the maximum cross-correlation between EMG channels and AU labels appeared. A similar method was used to calculate the delay between AU6 and AU12 from the CV-based binary labels.

Agreement with human coders. Human-coded labels, CV-labels, and EMG-labels were transformed to a matching sampling rate. Then the agreement between different measurements was calculated using Cohen's Kappa [4]. Additionally, we report accuracy, precision, and recall. The advantage of using Cohen's Kappa is that it penalizes for the larger amount of no AU samples in the set, given that participants did not smile all the time.

6 RESULTS

6.1 Delay between CV-EMG signals

The delay between raw EMG and CV-based labels was 374 ms in average (median = 450 ms, SD = 366 ms, Fig. 1).

6.2 Action Unit identification

6.2.1 Ground truth. The inter-coder agreement was a Cohen's Kappa of 0.78 for AU06 and 0.84 for AU12. For further processing, a single human-coded label was set to active when either of the coders thought there was an AU. Moreover, the agreement between CV and the human labeling was a Cohen's Kappa of 0.42 for AU06 and 0.43 for AU12.

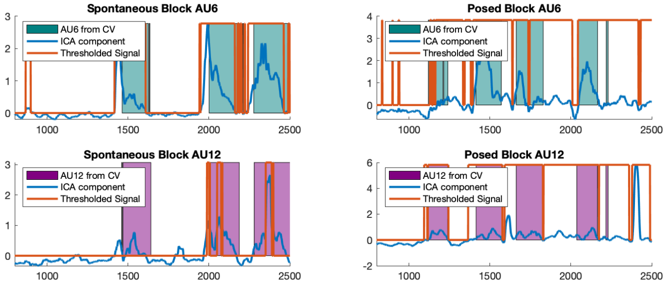

Figure 3: The threshold method applied to the ICA components identified as AU6 and AU12 for a block with spontaneous smiles (left) and posed smiles (right). The ICA components are show as blue lines.

Table 1: Agreement of AU6 and AU12 between human-coded and CV-based and between human-coded and CV-EMG detected AUs.

		Coder1 vs. Coder2	Coders vs. CV	Coders vs. EMG
AU6	Agreement κ	0.78	0.42	0.49
	Accuracy	-	0.83	0.81
	Precision	-	0.93	0.62
	Recall	-	0.34	0.60
AU12	Agreement κ	0.84	0.43	0.53
	Accuracy	-	0.83	0.82
	Precision	-	0.90	0.63
	Recall	-	0.35	0.68

6.2.2 Multimodal identification. The Cohen's kappa between Human-coded labels and EMG-based labels was 0.49 for AU6 and 0.53 for AU12. Furthermore, the accuracy reached 81% for AU6 and 82% for AU12 (Tab. 1). Although the correlation between CV-based labels and EMG-based labels is an important step for the selection of the EMG components, the selected EMG-based labels might not coincide perfectly with the CV-based labels (Fig. 3).

6.3 Synergy analysis

6.3.1 Co-occurrence analysis. To assess the co-occurrence frequency between AU6 and AU12, we evaluated how much agreement AU6 and AU12 labels have with each other. High agreement indicates that both AUs appear at the same time. For the human labeled data we see a Cohen's Kappa of 0.84 (Tab. 2). Similarly, the CV-based labels show a Cohen's Kappa of 0.62. Finally, the AUs detected from EMG show a Cohen's Kappa of 0.34. Moreover, a Wilcoxon rank sum test on the AU co-occurence pattern analysis (Fig. 4) showed that AU6 and AU12 co-occur more often than not (W = 6519.5, p < .01), regardless of the experimental block (W = 3362, p > .5). Furthermore, AU6 tends to start before AU12 most of the time (W = 7385, p < .01). An analogous analysis of CV vision data yielded

Table 2: Agreement of co-occurrence between AU6 and AU12 for human-coded, CV-based and from EMG detected AUs.

	Agreement κ	Accuracy	Precision	Recall
Human	0.84	0.96	0.77	0.77
CV	0.62	0.91	0.57	0.57
EMG	0.34	0.75	0.48	0.48

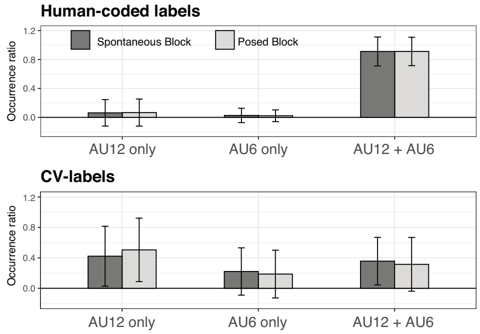

Figure 4: According to human coders, AU6 and AU12 co-occur most of the time. This differs from the labels according to OpenFace.

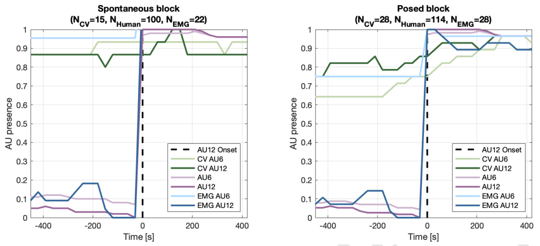

Figure 5: Frame-by-frame comparison of AU6 with respect to the onset of AU12 from human coded AUs, CV extracted AUs and AUs labeled with our method. The left figure shows results from spontaneous blocks, while the right figure shows the results from posed blocks. Human coded AUs are depicted in pink, CV extracted AUs in green and AUs from our method in blue.

no significant differences between posed and spontaneous blocks (W = 1390, p > .5), but showed opposite results regarding the order of occurrence between AU6 and AU12. According to CV, AU12 occurred before AU6 more often (W = 0.71, p < .001). To better understand the temporal relationship between both AUs, we performed a frame-by-frame comparison of AU6 and AU12 activation from human coded, with CV extracted AUs and AUs labeled with our method (Fig. 5). Each frame was selected with respect to the onset of AU12. A frame started 0.5 seconds before the onset and had a duration of 1 second. Whereas human coders rated an almost perfect co-occurrence, EMG and CV coded a higher probability of AU6 being active before AU12 onset.

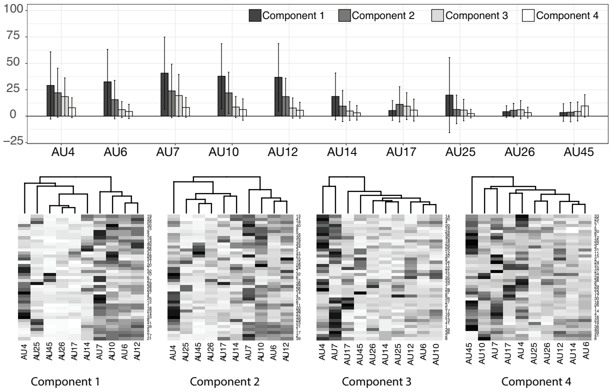

Figure 6: Weight averages per AU for the four components selected using the NNMF VAF as criteria. AU6 and 12 are grouped in components 1 and 2.

6.3.2 NNMF VAF Analysis on CV-based AU continuous labels. After iteratively decomposing the synergies from CV-AU labels using NNMF, the VAF analyses showed that four synergies account for 85% or more of the variance in both blocks for all participants A Wilcoxon rank sum test showed no differences in weights between spontaneous and posed blocks (W = 3968700, p > .05). A closer look to the weights of those four synergies showed that not all AUs have high weights. Therefore, we selected the AUs whose weights contributed more to the four synergies. The selected AUs are 4, 6, 7, 10, 12, 14, 17, 25, 26, 45 (Fig. 6). The resulting weights were clustered per participant and AU within each synergy. AU6 and 12 were grouped together in the first two components, which accounted for most of the variance. AU4 was often present, but not clustered with other AUs. In the third component, AU6 was clustered with AU10. Finally, blinks (AU45) emerged in the last component.

6.3.3 ICA-NNMF source-synergy analyis. Since Distal EMG is not measured close to the movement originating muscle, we proposed to use ICA first to approximate the movement sources before applying NNMF. After ICA, three sources were approximated and labelled using our matching algorithm. Next, only two NNMF components can be approximated due to the lose of one degree of freedom after ICA. The weights for the two resulting synergy components are shown in Fig. 7. Interestingly, the difference between experimental blocks of the synergy weights of AU6 and AU12 was significant according to a Wilcoxon rank sum test (W = 9886.5, p < .005). In the spontaneous block, one synergy is comprised of AU12 and others, and the second synergy is the joint activity of AU6 and AU12. On the contrary, the posed block yielded to two distinct synergies, one corresponding to AU6, and the other to AU12, which suggests that both facial movements are jointly executed only during spontaneous smiles. Finally, the "Other" label is comprised of any other AUs detected by the EMG electrodes.

6.3.4 Temporal matching. Fig. 8 shows the CV-derived components per AU that are most correlated to the EMG identified AU activations per experimental block. A Wilcoxon rank sum test showed no significant difference between posed and spontaneous experimental

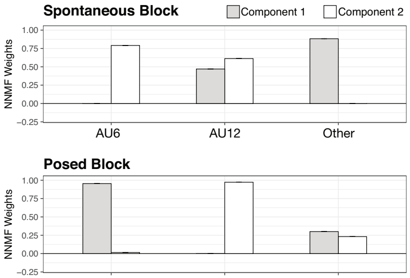

Figure 7: AU synergies identified per experimental block and NNMF-components. The horizontal axis shows the AUs detected by our matching algorithm. The vertical axis show the component weights per AU.

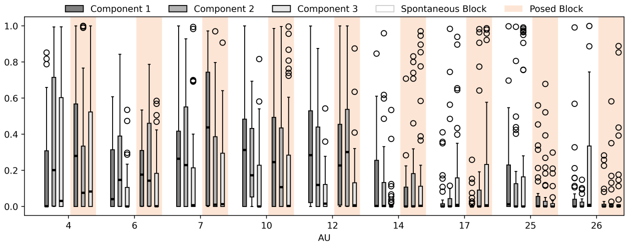

Figure 8: CV labeled AU weights after temporal matching with EMG components per experimental block. Components are grouped along the horizontal axis in AUs detected by OpenFace and experimental block. Each bar corresponds to the weight of a CV components with maximal correlation to the corresponding emg component.

blocks (W = 3968700, p > .05). Fig. 9 shows a heatmap depicting the synergy weights per AU and participants. The AUs with weight loads close to zero were excluded. Since no differences were found between experimental blocks, the weights among the two blocks are averaged. AU6 and 12 were clustered together in two components, whilst in the first AU6 was grouped with AU25, suggesting that the participants smiled with lips apart. Moreover, AU12 and AU10 were clustered together, which might be due to a confusion of the OpenFace algorithm.

6.3.5 Spatial matching. To match the dimensions of the EMG signal, which has four channels, four AUs were selected. These were AU6, AU7, AU10, and AU12. These are the AUs with the largest weight from previous analyses that are smile-related. Given the high occurrence of AU7 and AU10, we hypothesized that OpenFace might be confusing them with AU6 and AU12 respectively. A Wilcoxon rank sum test showed no differences in weights between spontaneous and posed blocks(W = 27872, p > .05). Figure 10 shows the synergy weights per AU. We can observe that AU6 and AU12 are clustered together in components 2 and 3. In component 1, AU6 is grouped with AU10 for most participants.

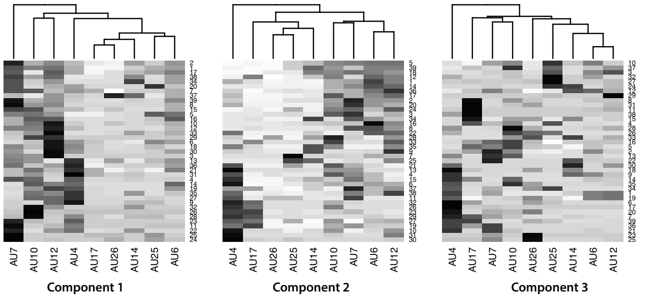

Figure 9: CV synergy cross-correlated with EMG IC Matching for selected AUs. Spontaneous and posed blocks were averaged given the lack of significant differences.

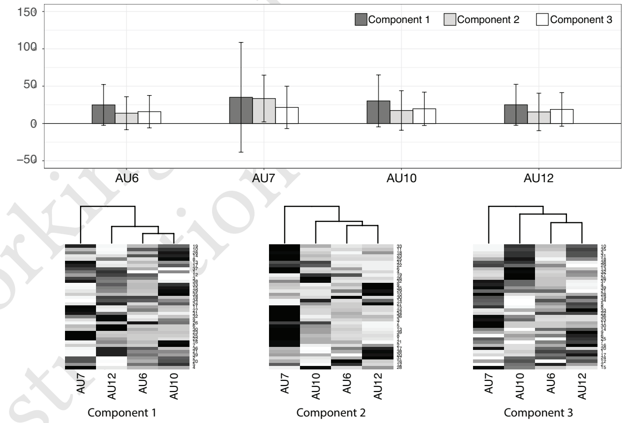

Figure 10: Four smile-related AUs were decomposed in three synergies for CV and EMG independently. The top plot shows the average weight loads per AU. The heatmaps below show the weight loads per participant and action unit per component. From the heatmaps, AU6 and AU12 were clustered together in components 2 and 3.

7 DISCUSSION

We observed co-varying activation patterns of pre-processed EMG, and continuous AUs extracted from CV. There was a delay between CV AU activation and EMG activation, with EMG activation leading by 374 ms. This was expected as EMG originates skin displacement, and it was observed before in proximal EMG measurements with an average of 230 ms [5].

Furthermore, we showed that AU6 and AU12 can be detected from a multimodal algorithm that labels EMG signatures estimated with ICA with labels generated automatically with CV. These EMG signatures do not always correspond to those dictated by the CV algorithm during the calibration. However, they yield to a slightly higher agreement than if CV was used alone. The main reason why the accuracy of the CV labels did not constrain the accuracy of the EMG-derived labels is that CV is used only for initial identification of the EMG components that are more correlated to each AU.

Moreover, we proposed the framework of *Sensing-Source-Synergies*. We distinguish the measurements made from the sources of the movement, and we use those sources to estimate synergies. We suggested to use ICA to search for activity from different muscle sources when measuring with EMG. On the other hand, NNMF attempts to find joint activation of different muscles. Several analysis pipelines in different sensing modalities were suggested to quantify the synergies between multiple sources. The human-coded co-occurring blocks of AU6 and AU12 showed that the Duchenne marker appeared simultaneously to the lip movements about 90% of the time, independently of the experimental block. Interestingly, AU6 leading AU12 occurred in more cases than the opposite. A similar analysis on the binary CV-based labels showed a lesser percentage of simultaneous activation, followed by an even less simultaneous activation depicted by the EMG labels. This is probably because the automatic algorithms tend to detect gaps in between AU6 activation when human coders do not report so. Nevertheless, in the cases where the Duchenne marker and the lip movent co-occurred, AU6 was active before AU12 above 80% of the time in the spontaneous block; and above 60% of the time in the posed block. The agreement of AU6 and AU12 labels per modality was the highest for human-coding, followed by CV and EMG. Probably EMG had the least co-occurence agreement given its higher temporal resolution.

To better interpret the amount of AU synergies, NNMF was also performed on the AU labels derived from CV, and a VAF metric was used to determine the ideal number of synergies. The results suggested four synergies. Moreover, the weights suggested that not all AUs contributed to these synergies, and no differences between posed and spontaneous experimental blocks was found. Another method proposed to select synergies was temporal matching with the EMG IC components. This method yielded three synergies with similar characteristics to the ones selected with the VAF criteria. Also, no significant differences were found between posed and spontaneous blocks. This entails that even though AU6 and AU12 weights are often clustered together within the selected synergies, there were no differences between posed and spontaneous smiles. According to these analyses, participants also displayed other AUs, notably AU4. This was expected, especially in the posed block where there might be a conflict between what the participants felt and the happiness they were asked to convey with a smile. It does not mean that those AUs are part of a smile, but that they were also present in the block. This suggested that antagonistic AU synergies might also occur, as to try to inhibit or mask facial expressions other than the intended one. Furthermore, we observed that OpenFace often lead to high weights in AUs that are similar to AU 6 and 12. For example, AU6 might be confused with AU7, and AU12 with AU10. Thus, when we had to further reduce the AUs to match to the EMG results, we chose these. The results showed that AU6 and 12 are often grouped together, and in one case, AU6 was grouped with AU10. Thus, whilst CV-based identification has a higher spatial resolution to identify several AUs, it still confuses several labels. Moreover, NNMF was used to identify joint muscle activity directly from the pre-processed EMG in an spatial matching algorithm. The results were similar to those found by applying NNMF on CV AUs only. This is in line with the hypothesis that NNMF considers the ZM and the OO to move as one single synergy. A closer look into the NNMF results suggested that the aforementioned muscles move in a single synergy picked up by the lower electrodes of the wearable. Thus, probably the synergy is dominated by the ZM. This is in line with results showing the significant strength of the ZM when compared to other muscles, and it might be related to muscle length [41].

We also used the results of our proposed multimodal identification of AUs using EMG, as an input to the NNMF algorithm. Although the number of synergies that can be identified using this method is only two, the results were surprising. We had hypothesized that a challenge for ICA might be to disentangle multiple AUs from the mixed EMG given the joint activation of the ZM and the OO. However, the results showed that ICA actually boosted the NNMF synergy detection by transforming the data to the source space first. Only in this case, we found a strong difference between posed and spontaneous smile AU synergies. In posed smiles AU6 and 12 seem to operate independently, whereas a joint movement of AU 6 and 12 was observed in spontaneous smiles. This is somehow in line with the Duchenne marker hypothesis. Even though the appearance of the Duchenne marker can be simulated voluntarily, the underlying muscle synergies are distinct. The success of EMG to recognize subtle differences might be that (1) with our matching algorithm EMG already contains information from the CV-based labeling; (2) the forced use of a reduced set of synergies might have made the differences more salient.

8 LIMITATIONS OF THIS STUDY

One of the limitations of this study was the number of electrodes provided in the EMG wearable. Whilst four electrodes provide a good trade-off between wearability, smile and AU detection, they are limited to conduct muscle synergy analysis. Therefore, we opted to determine the optimal number of synergies using CV only, and a synergy-matching strategy between EMG and CV. Increasing the electrode number will enable us to explore synergies containing more facial expressions. In this case, we opted to model mainly AU6 and AU12, and to consider other AUs in the EMG as "noise". Moreover, synergy analysis based on NNMF requires continuous labels. The human-labeled AUs were performed frame-by-frame, but the coders only indicated presence, not intensity, thus, we opted to do a block analysis on single and simultaneous activation. Labeling AU intensity as well would have been useful to apply our NNMF method directly on the ground-truth labels. Fortunately, OpenFace derives both continuous and binary AU labels which were useful for our method. However, the CV-only detection still could be improved. In particular, we found out that CV might have confused AU10 for AU12; and AU6 for AU7. Finally, we conducted this study on data aimed to elicit smiles. It would be interesting to assess the synergies present in other types of facial expressions.

9 CONCLUSIONS AND FUTURE DIRECTIONS

Our Sensing-Source-Synergy approach led to good AU identification from a multimodal system, and aided to fine-grained AU synergy analysis. Our results suggest that the Duchenne marker can be displayed in both posed and spontaneous smiles, but the underlying synergy differs. Moreover, in recording session where high movement or high facial occlusion are expected, CV alone would struggle to continuously identify certain AUs. On the other hand, wearable distal EMG can deal with occlusion and movement, but it cannot disentangle AUs so easily. By combining both and using CV in a calibration period as a reference for automatic AU identification in EMG as presented here, we would be able to create a more robust system that enables fine-grained analysis of facial expression synergies and activation. A use case scenario would be when assessing children's behavior. It would be possible to have a short calibration session with both camera and EMG recordings. Using this multimodal information, the EMG could be tagged and used to identify AUs while the children play and run around, without requiring constant camera surveillance.

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. https://doi.org/10.1109/FG.2018.00019

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

[3] Ho Seung Cha, Seong Jun Choi, and Chang Hwan Im. 2020. Real-time recognition of facial expressions using facial electromyograms recorded around the eyes for social virtual reality applications. *IEEE Access* 8 (2020), 62065–62075. https://doi.org/10.1109/ACCESS.2020.2983608

[4] Jacob Cohen. 1968. Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychological Bulletin* 70, 4 (1968), 213–220. https://doi.org/10.1037/h0026256

[5] 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. https://doi.org/10.1142/S021969130400041X

[6] Pierre Comon. 1994. Independent component analysis, A new concept? *Signal Processing* 36, 36 (1994), 28–314.

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

[8] Paul Ekman. 1999. Basic Emotions. In *Handbook of cognition and emotion*, Tim Dalgleish and M. Power (Eds.). John Wiley & Sons, Ltd., Chapter 3, 45–60. https://doi.org/10.1017/S0140525X0800349X

[9] Paul Ekman, Wallace Friese, and Richard Davidson. 1988. The Duchenne Smile: Emotional Expression And Brain Physiology II. *Journal of Personality and Social Psychology* 58, 2 (1988), 342–353.

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

[11] Paul Ekman and Wallace P. Friesen. 1982. Measuring facial movement with the Facial Action Coding System. In *Emotion in the human face* (second edi ed.), Paul Ekman (Ed.). Cambridge University Press, Chapter 9, 178–211.

[12] Paul Ekman, Wallace V. Friesen, and Maureen O'Sullivan. 1988. Smiles when lying. *Journal of Personality and Social Psychology* 54, 3 (1988), 414–420. https://doi.org/10.1037/0022-3514.54.3.414

[13] 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. https://doi.org/10.1007/s10803-013-1898-4

[14] Jeffrey M. Girard, Gayatri Shandar, Zhun Liu, Jeffrey F Cohn, Lijun Yin, and Louis-Philippe Morency. 2019. Reconsidering the Duchenne Smile: Indicator of Positive Emotion or Artifact of Smile Intensity? *PsyArXiv* (2019). https://doi.org/10.31234/OSF.IO/Z2JVD

[15] 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. https://doi.org/10.1109/TAFFC.2014.2313557

[16] 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. https://doi.org/10.3389/fpsyg.2018.00202

[17] Mahyar Hamedi, Sh-Hussain Salleh, Mehdi Astaraki, and Alias Mohd Noor. 2013. EMG-based facial gesture recognition through versatile elliptic basis function neural network. *Biomedical engineering online* 12, 1 (2013), 73. https://doi.org/10.1186/1475-925X-12-73

[18] Muhammad Haris Khan, John McDonagh, and Georgios Tzimiropoulos. 2017. *Synergy between face alignment and tracking via Discriminative Global Consensus Optimization*. Technical Report. 3791–3799 pages.

[19] Rayner Pailus Henry and Rayner Alfred. 2018. Synergy in Facial Recognition Extraction Methods and Recognition Algorithms. In *Lecture Notes in Electrical Engineering*, Vol. 488. Springer Verlag, 358–369. https://doi.org/10.1007/978-981-10-8276-4_34

[20] 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

[21] Lilah Inzelberg, Moshe David-Pur, Eyal Gur, and Yael Hanein. 2020. Multi-channel electromyography-based mapping of spontaneous smiles - IOPscience. *Journal of Neural Engineering* 17, 2 (apr 2020). https://iopscience.iop.org/article/10.1088/1741-2552/ab7c18/meta

[22] Martin Köstinger, Peter M. Roth, and Horst Bischof. 2012. Synergy-based learning of facial identity. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Vol. 7476 LNCS. Springer, Berlin, Heidelberg, 195–204. https://doi.org/10.1007/978-3-642-32717-9_20

[23] Eva G. Krumhuber and Antony S. R. Manstead. 2009. Can Duchenne smiles be feigned? New evidence on felt and false smiles. *Emotion* 9, 6 (2009), 807–820. https://doi.org/10.1037/a0017844

[24] Daniel D. Lee and H. Sebastian Seung. 1999. Learning the parts of objects by non-negative matrix factorization. *Nature* 401, 6755 (oct 1999), 788–791. https://doi.org/10.1038/44565

[25] Brais Martinez, Michel F. Valstar, Bihan Jiang, and Maja Pantic. 2017. Automatic Analysis of Facial Actions: A Survey. *IEEE Transactions on Affective Computing* (jul 2017). https://doi.org/10.1109/TAFFC.2017.2731763

[26] Daniel S. Messinger. 2002. Positive and negative: Infant facial expressions and emotions. *Current Directions in Psychological Science* 11, 1 (feb 2002), 1–6. https://doi.org/10.1111/1467-8721.00156

[27] 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. https://doi.org/10.1007/s12144-016-9448-9

[28] 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 (sep 2007), 167–78. https://doi.org/10.1080/17470910701391943

[29] Monica Perusquía-Hernández, Saho Ayabe-Kanamura, and Kenji Suzuki. 2019. Human perception and biosignal-based identification of posed and spontaneous smiles. *PLOS ONE* 14, 12 (dec 2019), e0226328. https://doi.org/10.1371/journal.pone.0226328

[30] 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 *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*. ACM Press, New York, New York, USA, 1–9. https://doi.org/10.1145/3290605.3300379

[31] 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. https://doi.org/10.1109/TAFFC.2017.2755040

[32] 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.

[33] 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. https://doi.org/10.1007/s10919-008-0058-6.Comparison

[34] 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.

[35] IT Schultz and Martin Pruzinec. 2010. *Facial Expression Recognition using Surface Electromyography*. Ph.D. Dissertation. Karlruhe Institute of Technology. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.188.1435{&}rep=rep1{&}type=pdf

[36] 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. https://doi.org/10.1145/2593968.2610455

[37] Chun Kwang Tan, Hideki Kadone, Hiroki Watanabe, Aiki Marushima, Masashi Yamazaki, Yoshiyuki Sankai, and Kenji Suzuki. 2018. Lateral Symmetry of Synergies in Lower Limb Muscles of Acute Post-stroke Patients After Robotic Intervention. *Frontiers in Neuroscience* 12, APR (apr 2018), 276. https://doi.org/10.3389/fnins.2018.00276

[38] Gelsy Torres-Oviedo and Lena H. Ting. 2007. Muscle synergies characterizing human postural responses. *Journal of Neurophysiology* 98, 4 (oct 2007), 2144–2156. https://doi.org/10.1152/jn.01360.2006

[39] Gelsy Torres-Oviedo and Lena H. Ting. 2010. Subject-specific muscle synergies in human balance control are consistent across different biomechanical contexts. *Journal of Neurophysiology* 103, 6 (jun 2010), 3084–3098. https://doi.org/10.1152/jn.00960.2009

[40] Matthew C. Tresch and Anthony Jarc. 2009. The case for and against muscle synergies. , 601–607 pages. https://doi.org/10.1016/j.conb.2009.09.002

[41] 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.

[42] 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

[43] 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.