# Dynamics of Positive Affective Responses Identified through Behavioral and Electrophysiological Measures

March 2018

Perusquía Hernández Monica

# Dynamics of Positive Affective Responses Identified through Behavioral and Electrophysiological Measures

School of Integrative and Global Majors
Ph.D. Program in Empowerment Informatics
University of Tsukuba

March 2018

Perusquía Hernández Monica

# Dynamics of Positive Affective Responses Identified through Behavioral and Electrophysiological Measures

**Monica Perusquía-Hernández**

**Abstract**

Facial expressions are among the most salient cues for automatic emotion identification. However, they do not always represent felt affect. As they are an indispensable social communication tool, they can also be fabricated to face complex situations in social interaction. In this work, identification of spontaneous and posed smiles was explored using multimodal wearable sensors. Distal facial Electromyography (EMG) can be used to differentiate between them robustly, unobtrusively, and with good temporal resolution. Furthermore, such wearable can be enhanced with autonomic electrophysiological and behavioral signals. Electrodermal Activity (EDA)-based recognition is a good indicator of affective arousal. Head movement is useful to capture the inherent movement during spontaneous expressions. In three experiments, six research questions were asked and answered. In Experiment 1, distal EMG was shown as an effective measure to identify fast and subtle spontaneous smiles, even at a micro-expression level. This is specially useful when two or more people are being tracked. Moreover, it was possible to identify the differences between posed and spontaneous smiles. Whilst the spatial distribution of the muscles differs, temporal features are more robust to distinguish among them. Namely, smile duration, rising time, and decaying time. Experiment 2 confirmed the potential of using EMG to identify the smile's spatio-temporal dynamics. Special care was taken to elicit posed smiles intended to convey happiness. In this case, rising time and decaying speed differed significantly. EDA and IMU measures alone also have the potential to distinguish between co-occurring spontaneous and posed smiles with high accuracy. IMU-measured data explained best their differences. Moreover, no cultural differences were found between posed and spontaneous smiles from their embodied measures. These results seem to support the view that embodied affective responses are similar for all humans, regardless of their cultural background. Observed behavior was clearly related to the self-reported measures in both experiments. Experiment 3 showed that laypersons can distinguish between posed and spontaneous smiles above chance level with modest accuracy. Hence, using behavioral and electrophysiological signals complements human ability, as they provide information not visually perceivable. Even though psychologists have asked similar questions in the past, none has addressed them with a multimodal wearable. This approach is a promising tool to explore with more temporal resolution how emotion processes arise and develop in our bodies. Moreover, this work has only proven the potential of the proposed approach. Future work should consider improving the wearable system for comfortable real-time usage in more ecologically valid settings.

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

Since the moment we are born, we experience and express emotion. We cry for help; we express joy to create bonding; we experience fear as a survival cue; and we experience shame as a self-protection mechanism. Emotion plays a central role in our lives, as it changes the way we perceive ourselves; how we make decisions; and how we interact with others.

Affect and emotion are central to human experience. All choices we make are based on what makes us feel good, what improves our overall experience. This applies to personal relations, products, and services. Therefore, assessing human experience is relevant in many application domains. These range from historical documentation; tracking therapy results and augmented feedback for impaired people[1, 2]; user and customer experience mapping[3–5]; human-robot interaction[6]; and even personal management.

Specially in marketing and design, positive user experience is an important metric of the success of a product, media, or service. 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 again to check for improvements[7, 8].

Despite its importance, emotion is among the most subjective topics of discussion. Everybody has experienced it, but in a personal, private manner. This experience

is subjective, and prone to several biases. Hence, it has always been a challenge to quantify it in an efficient and objective fashion.

This assessment is usually done via qualitative methods like interviews; or quantitative methods such as tools of explicit and implicit self-report. For example, Csikszentmihalyi argued that to assess dynamics of mental health and user experience it is "essential to develop measures for the frequency and the patterning of mental processes in every-day-life situations" [9]. As a solution, he proposed his renown Experience Sampling Method (ESM) [10] to provide a valid method to describe variations in self-reports of mental processes. These mental processes include frequency and intensity of psychological states such as emotional, cognitive, and conative dimensions of user experience. For longitudinal studies, this method provides a good approximation of what the user is feeling along with the use of a product. However, it requires a logging tool that intermittently prompts users to report their experience.

Another challenge of using self-report measures is the multiple biases that people have when answering them. For example, the Social Desirability bias, and the Hawthorne Effect bias. The Social Desirability Bias refers to the fact that people tend to self-report inaccurately to present themselves in the best possible light [11]. The Hawthorne Effect describes the fact that people tend to behave differently because they are aware of being observed [12]. These biases are more salient when using qualitative methods. Furthermore, analyzing the data while carefully trying to minimize those biases is a time-consuming process.

On the other hand, psychologists have developed several measurement tools that try to reduce these. For example, the Affective Grid developed by Russell [13], the Self-Assessment Manikin (SAM) [14], the Affective Slider [15]; or the Implicit Positive and Negative Affect Test (IPANAT) [16] and the Inkblot test [17] for implicit self-report. Even though these tools are validated in large samples of people to ensure their reliability, there are still some points of improvement. First, they require time from the user to fill them in. Second, it is unpractical to fill them multiple times to assess experience continuously. This would interrupt the experience itself. Third, the

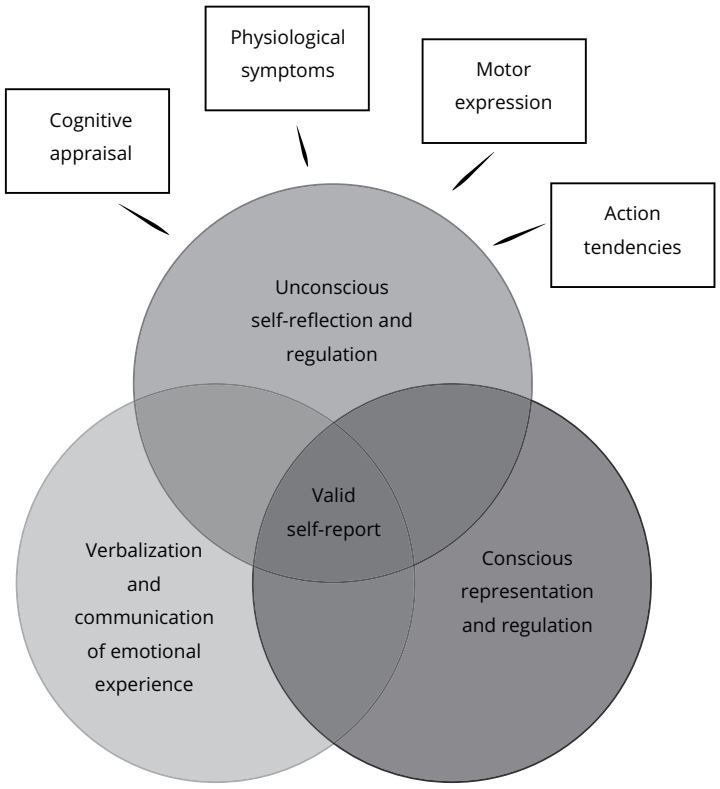



Figure 1.1: *Valid self-report zone.* Three modes of representing changes in emotion components are suggested by Scherer. Unconscious, conscious, and verbal. The zone of valid self-report is the overlap among these.

temporal resolution in which we can sample user experience with self-reported tools is limited. This might be because humans are not aware of their emotional state at every instant. Previous studies have shown that awareness rates of facial expressions are about 60%. Awareness depends on intensity and (or) duration of the facial expression [18]. Scherer (2005) pointed out that, given the complexity of emotion, there are three modes of representing the changes in emotion components: unconscious, conscious, and verbal. The zone of valid self-report is the overlap among these three components (figure 3.1). Therefore, there are physiological symptoms, motor expressions, action tendencies and cognitive appraisals that cannot be fully assessed through self-report [19].

Recently, a popular alternative to assess human affect in a continuous manner is to measure and interpret behavior and electrophysiological cues automatically us-

ing Artificial Intelligence (AI) technology. Computer Vision (CV) has been used to identify facial expressions or posture. Other sensors have been used to measure Autonomic Body Responses such as Heart Rate Variability (HRV), Electro-Dermal Activity (EDA), and Electromyography (EMG). Several surveys have been done to summarize the different signals that can be used for technology technology-afforded emotion recognition [20–22]. However, due to the plethora of experimental paradigms, signal type, features, and classification schemes to identify different combinations of emotions; it is very difficult to compare them to choose the best combination. Additionally, the number of studies using multiple modalities is limited, compared to unimodal studies. Therefore, there is still room for improvement in multimodal-based emotion recognition, specially regarding data fusion at different levels, i.e., feature, classifier, or model.

Identifying an affective state using technology is thus possible and has several advantages for continuous assessment of affective experience. The main advantage is that they have the potential to provide an uninterrupted reliable measurement with high temporal resolution. However, an important step to identify an affective state using these measures is to define the relationship between the self-reported emotion label, and its correspondent embodied response. There are different methods of establishing the so-called ground-truth. Video-coding of facial expressions, self-reported labels, and most importantly, labels according to how the data was acquired.

Furthermore, the ground truth label itself depends on the predefined labels for affective states that are chosen beforehand. For centuries psychologists have debated on what are the universal emotions and how affect processes occur, and how they can be represented. These include the theory of basic emotions [23], dimensional theories of emotion [24], and appraisal theories of emotion [25].

Moreover, the emotion labels and its physical expression might differ among cultures. Previous research has shown some evidence that basic emotions have universal facial expressions [26]. However, this universality view is still under debate. The observer of a facial expression might interpret this behavior according to a different

set of labels than the predefined by the researchers; interpret it according to a situation as a part of an instrumental action; or as a group of facial sub expressions. It is also possible that the facial expressions might be interpreted in terms of bipolar dimensions rather than discrete labels [27]. All in all, these alternative interpretations might be somehow regulated through learning and cultural context.

Among the different representations of affect, the Russell's Circumplex model of affection seems most suitable for continuous measurement because it represents multiple emotion labels in two dimensions. These dimensions are valence and arousal. Valence and arousal can be mapped to embodied affect responses. Facial expression detection (video or electromyography -based) is a good predictor of the valence an emotion, and physiological signals (e.g., Galvanic Skin Response) are a good representative of the arousal of an emotion. Furthermore, behavioral cues such as head and body movement also carry information about the affective state of the person. These can be combined in a wearable which can provide information about facial expressions and their nature.

Facial expressions can represent both positive and negative affect. Previous research showed that it is possible to detect facial expressions with both CV [28] and EMG measurements[29, 30]. CV has a good spatial resolution, is unobtrusive, and can distinguish between several facial expressions. However, it has important limitations to detect fast and subtle facial expressions, or multiple users at the same time. On the other hand, EMG was shown to be able to detect both smiles, i.e., positive valence; and frowns, i.e., negative valence [30]. This approach has important advantages. First, EMG has a high temporal resolution, which is suitable to detect fast changes. Second, since the EMG is measured distally, it does not cover the user's face, and signals from different muscles can be read. However, its spatial resolution is not as good as CV, and its advantage measuring fast and subtle expressions or multiple persons at the same time is yet to be tested.

As promising as it seems, to measure user experience, detecting and counting smiles and frowns is not enough. Facial expressions can be spontaneous affective re-

sponse to a stimulus; but they can also be a posed response to a cognitive intention. On one hand, facial expressions are usually linked to emotional states of a person, and they are among the most salient cues of the experienced emotion [26]. On the other, these facial expressions can also be voluntarily fabricated, which indicates that similar facial expressions do not always accurately describe the actual experienced emotion [31, 32]. According to [33], "the movements inherent to posed facial expressions display an emotion an expresser ostensibly intends to convey, whereas spontaneous facial expressions correspond to an expresser's actual, unmitigated emotional experiences." Because a spontaneous facial expression is an automatic motor movement, and posed facial expressions are voluntary, they are also believed to have different neural pathways [34], and they are represented by different temporal dynamics from EEG signals [35, 36]. Furthermore, some researchers have also stated that there is a difference between facial expressions which have a communicative intent in social contexts, and expressions without social intent, adding another dimension of analysis.

Perhaps the most commonly studied facial expression regarding posed and spontaneous differences is a smile. Besides expressing happiness, a smile can also be used to convey kindness to others. Different terms have been used to refer to these smile-types. Posed and deliberate smiles are often used as synonymous, and are opposite from spontaneous smiles. Several differences among posed and spontaneous smiles have been found. The most sound difference is the activation of the *orbicularis oculi* muscle that was believed to happen during spontaneous smiles only [37]; in the so-called Duchenne smile. Nevertheless, recent studies have found that this muscle is activated both in posed and spontaneous smiles [33]. Furthermore, these smiles have been found to differ in amplitude [34, 38, 39], in their temporal dynamics [34, 38–42], and in the behavioral movements that accompany them. The most salient movement being the head movement [43]. Most of this evidence has been found sing Human Coding and CV methods. However, these have the aforementioned limitations. In contrast, physiological signals have a higher temporal resolution, and they are able to pick up information even if it is not visually perceivable.

Alongside with the valence response changes measurable through facial expressions, arousal changes can be measured using EDA. Several studies have explored how skin conductance varies along responses to affective stimuli [44]. Moreover, as fast autonomic changes cannot be controlled voluntarily, they might prove to be a good cue to distinguish between spontaneous and polite responses.

It is proposed to use multimodal wearable technology to assess the quality of the affective experience of a person. Since embodied affective responses can be controlled voluntarily, only quantizing these cues is not enough. It is also necessary to distinguish between spontaneous reactions and voluntary expressions. Therefore, it is proposed to use EMG to detect facial expressions [30], EDA to judge emotional arousal (i.e., intensity) [44]), and head orientation via an IMU to further differentiate posed and spontaneous responses [43]. As a proof of concept, the focus is mainly on positive affective responses. In other words, responses occurring during posed and spontaneous smiles. Positive affective responses were chosen due to their social relevance. The proposed wearable approach has the advantage of providing an unobtrusive log of user experience. Furthermore, the behavioral and electrophysiological measures chosen can be sampled at high temporal resolution, allowing for a fine-grained analysis of the spatio-temporal dynamics of different affective responses.

## 1.1 Research questions and thesis outline

As previously motivated, the feasibility of continuously logging user experience using electrophysiological and behavioral measures is investigated. For this purpose, identifying the difference in the dynamics of posed and spontaneous affective responses is of utmost importance. By distinguishing among them, affective experience can be better assessed. The specific Research Questions (RQ) to be answered during this research are the following:

- **RQ 1.** Is it feasible to use EMG to detect positive fast and subtle facial expressions at the micro-expression level?

- **RQ 1-b.** Is it feasible to detect positive fast and subtle expressions through EMG from multiple users at the same time?
- **RQ 2.** Can we distinguish a positive affective posed reaction from a spontaneous one using EMG?
- **RQ 2-b**. What are the differences in EMG spatio-temporal dynamics of posed and spontaneous smiles?
- **RQ 3.** How does a multimodal system, including head movement and EDA, improve the identification of posed and spontaneous positive affective responses?
- **RQ 4.** How does the observed behavior occurrence and dynamics relate to self-reported measures of affect?
- **RQ 4-b.** Is there any difference in that relation between implicit and explicit self-reports?
- **RQ 5.** How does cultural background affect our affective responses regarding the investigated measures?
- **RQ 6.** How good are humans at distinguishing between posed and spontaneous smiles?
- **RQ 6-b.** What are the advantages of identifying spontaneous and posed smiles using behavioral and electrophysiological data versus other methods?

To investigate these questions, three main experiments were conducted. The main purpose of the first two experiments was to collect affective response data to positive stimuli.

The first experiment served to collect spontaneous and posed facial expressions from EMG, in particular, micro-expressions. In a second version of experiment 1, data from pairs of participants was obtained simultaneously. The first experiment addressed research questions 1 and 2.

The second experiment focused in eliciting posed and spontaneous smiles from participants of multiple cultural backgrounds whilst taking multiple affect measures. These measures included facial EMG, EDA from multiple locations, Heart Rate, video, video coding of own facial expressions, and explicit and implicit self-reported measures. The second experiment addressed research questions 2 till 5, and partially RQ6.

The third experiment's purpose was to investigate RQ6, by using data from experiment 2 as stimuli.

This thesis is structured as follows. In Chapter 2, a detailed background on affective responses is outlined. Chapter 3 describes in detail the challenge of measuring affective responses, the different measurement tools available, and their advantages and disadvantages. Chapter 4 introduces experiment 1 on a wearable device for fast and subtle smile recognition, followed by the data analysis on the spontaneous and posed smile recognition based on spatial and temporal patterns of facial EMG. Chapter 5 introduces experiment 2. From this data, further differences from spontaneous and posed smiles from EMG are investigated. Moreover, the relationship between these facial expressions, behavioral and autonomic affective reactions, self-report, and cultural differences was explored. Chapter 6 describes the relationship between self-report and the measured embodied responses using data from both experiments. Chapter 7 describes experiment 3 on the human judgment of posed and spontaneous smiles. Chapter 8 includes a general discussion, and outlines potential applications of this research. Finally, Chapter 9 describes conclusions and future directions.

# Chapter 2

# Affective responses

## 2.1 Affection and emotion

Affective phenomena (emotions, moods, and affect) have been explored extensively in psychology over more than a century. According to Ekkekakis [45], core affect is a neurophysiological state accessible as a simple primitive non-reflective feeling, and it is part of both emotion and mood. Moods are long-lasting affective episodes that do not have an apparent cause. They are about nothing specific or about everything. In contrast to moods, which typically last longer, emotion is a short-term affective reaction to an object, agent, or event [20].

Emotions are complex affective events that consists of a core affect; an overt behavior congruent with the emotion; attention directed towards the eliciting stimulus; attribution of the genesis of the episode of the stimulus; the experience of the particular emotion; and neural and endocrine changes consistent with the particular emotion [45].

Different theories have been proposed on how emotion processes start, and about their relation to embodied cues and cognition. Early proposals consider that emotions are bottom-up processes. In other words, changes in our bodily states create the subjective feeling of an emotion. In 1872, Darwin had already pointed out the evolutionary function of emotion [46]. According to him, emotions solve certain problems

that humans face as a species. For example, a baby might cry to call for help. Humans share a common pattern of emotional expression via facial expressions as result of natural selection [47]. Therefore, the emotions conveyed by the face are universal [46]. In accordance with this point of view, the James-Lange Theory estates that embodied changes follow directly the stimuli that caused the emotion. Hence, the subjective feeling arises from that embodied state [48]. Basically, this point of view states that we are afraid of a wild animal because our body instinctively flees. Similarly, we are sad because we cry. Thus, this theory supports the view that emotions can be differentiated by somato-visceral responses. The Somatic Marker Hypothesis by Damasio further supported this point of view. He argued that there are marker signals that influence the processes of response to stimuli at multiple levels. Some of these occur consciously, and others are covert and occur non-consciously. He argued that those markers are somatic, as they arise in the brain's representation of the body [49,50]. In the same line of thought, Zajonc argued that affect precedes cognition and it does not require prior cognitive appraisal [51].

On the other hand, other theorists have described the influence of top-down cognitive processes in emotion generation. Cannon revised the James-Lange theory, and argued that emotions are derived from subcortical centers. This could explain why emotions can be elicited directly from brain stimulation [52]. Furthermore, dimensional appraisal theories state that all emotions are generated due to appraisal judgments [53, 54]. A surprise can be appraised positively or negatively according to the context. Therefore, changes in facial expressions are appraisal driven [54]. The Componential Emotion Theory states that subjective feelings emerge when the synchronization or coherence of appraisal-driven changes between emotion components has reached a critical threshold. In this Component Process Model (CPM), coherent response changes (i.e., appraisal, facial expressions, physiological changes, action tendencies, and subjective feeling) emerge despite the different response dynamics (i.e., latency, patterning, and intensity) in each emotion component during an emotional episode. In contrast with bottom-up theories, appraisals are assumed to

initiate the response changes in the peripheral components. Appraisals usually check the relevance of the stimulus, the implications for the personal well-being, the coping potential of the individual to face the event, and the normative significance of the event for the values of the individual. In this sense, the same stimulus might lead to different appraisals by different individuals, and therefore cause the subjective and private experience of emotion [25].

These theories consider the processes that are involved in emotion generation. On the other hand, there is also a discussion among psychologists on how emotions should be represented. The most accepted representations are categorical and dimensional representations. The Theory of Basic Emotions [23], is one of the earliest attempts to define and classify emotions. It proposes that there is a set of basic emotions universal to all cultural groups. These emotions are represented according to a label dependent on the language used. Different theorists have proposed several emotions. Whereas James [48] proposed fear, grief, love and rage as basic emotions, the most widely used set is the one proposed by Ekman [55]. In his view, there are six basic emotions: anger, disgust, fear, joy, sadness, and surprise. These six emotions are universally expressed by a emotion-specific physiology, and distinctive universal signs [26]. These signs include a set of prototypical facial expressions. This is an evolutionary view inspired by the one proposed by Darwin [32].

A drawback of the Theory of Basic Emotions is that this representation is constrained by the language used to express the labels for each emotion. In [27], it is argued that facial expressions and emotion labels are probably associated, but the association may vary per culture, and it is unable to explain nuances that are not included in the vocabulary used.

The dimensional theory of emotion aims to overcome this limitation by reducing the complexity of the representation of emotions. Several discrete emotion labels have been mapped to a limited number of dimensions. Perhaps the most popular is the circumplex model of emotion by Russell [24]. This suggests that emotions are distributed over a two-dimensional plane. These two dimensions are orthogonal to each

other. The first one is arousal, which ranges from inactive to active. The second is valence, which ranges from negative to positive. Some researchers have also suggested the use of other dimensions. A common one is a third axis, dominance, ranging from in control, to dominated [14]. This dimension allows to clearly differentiate discrete labels that are very close to each other in the two-dimensional space of the circumplex model. Other authors have also proposed four dimensional spaces, with evaluation-pleasantness, potency-control, activation-arousal, and unpredictability as dimensions [56].

All in all, emotion is a complex phenomenon. Simple representations allow scientists to approximate measurements of its multiple dimensios. However, as more knowledge is gained, emotions should be considered as a complex phenomenon that involves (a) appraisals of events, (b) psychophysiological changes, (c) motor expressions, (d) action tendencies, (e) subjective experiences, and (f) emotion regulation [56]. In this thesis, the relationship between psychophysiological changes and self-reported experiences during different motor expressions are explored. Particularly to this study, it is interesting to explore the relationship between similar facial expression and the presence or absence of a felt emotion. Furthermore, the changes in other physical responses during these expressions of emotion are explored. Scherer [19] argued that there are three modes of the representation of changes in emotion components. These are: (1) unconscious reflection and regulation, which includes physiological symptoms, motor expressions, cognitive appraisals, and action tendencies; (2) conscious representation and regulation; and (3) verbalization and communication of emotional experience. According to this, the zone of valid self-report report measurement is the overlap between those three (figure 3.1). This suggests that emotion processes are largely hidden to the person experiencing them, and it may be an explanation why emotion is often regarded as involuntary or unreasonable [25].

## 2.2 Affective embodiment

As pointed out by several of the aforementioned theories, affective states have an embodied manifestation. Both bottom-up and top-down theories acknowledge the role of the body in the emotion processes. As Prinz [57] proposed in the Perceptual Theory of Emotion, emotions are embodied appraisals. They are perceptions of the body, but, through the body, they also allow us to perceive several concerns, such as danger and loss. Appraisal supporters further suggests that appraisals are a type of perception, which enables us to feel, but also limits our ability to control the process. Moreover, motivational input into the appraisal process provides another opportunity for the appraisal to be involuntary, and thus, appear irrational [25].

Affective embodied responses include changes in the autonomous nervous system, facial expressions, or behavioral changes such as the degree in which we move our body. Some of those embodied responses are private, whereas others can be used as a communication tool to show others our affective estate.

Facial expression studies are perhaps the most studied affective responses. As mentioned before, several scientists have studied them extensively. Several facial expressions are believed to be hardwired and mapped to a specific felt emotion. Evidence from congenitally blind people who smile when happy, or display sadness supports this view [58]. However, these can also be used to provide misleading information about the wearer's emotional state [23, 31, 32, 37, 59]. The information conveyed by facial expressions is sent not only to third persons, but also to oneself. The facial feedback hypothesis [60] states that the feedback from the facial muscles is important for the subjective experience of emotion, in concordance with most of the bottom-up theories. However, other studies have argued that we are aware of our own facial expressions only after a certain threshold has been crossed. In this case, the intensity and the duration of the facial expression are important factors for awareness [18].

Autonomic changes and its occurrence associated with emotion have also been researched over the past century. Although it is widely accepted that these two phenomena co-occur, it has been stated that it is not possible to find unique and invariant

autonomic signatures to emotion. Autonomic responses are associated mostly to dimensional representations. Moreover, negative emotions are usually associated with more prominent autonomic responses than positive emotions. Furthermore, the autonomic activity is regarded as behavior preparation, and it is expected to occur before the behavior has been initiated [61].

Autonomic responses include Electro-Dermal Activity (EDA) or Skin-conductance (GSR), and Heart Rate Variability (HRV). EDA is consistently regarded as an indicative of cognitively or emotionally mediated motor preparation [61]. Therefore, an increase of EDA is observed in emotions other than non-crying sadness, acute sadness, contentment, and relief [61]. However, no specific consensus has been found from previous research. Whereas a study found support for the facial feedback hypothesis by measuring increased EDA during amplification of facial expressions than during inhibition, another group found the opposite effect [62].

Most of these changes are reflexes that happen automatically as response to an external or internal (i.e., an appraisal) stimuli. It is still under debate whether the physical response precedes cognitive appraisal or vice-versa. However, the fact remains: there are bodily states that covariate together with self-reported affective states, suggesting that these affective reactions provide an embodied private experience. Interestingly, embodied affective responses change in the order of minutes, seconds, or even milliseconds. On the other hand, people usually report affective changes in the order of minutes or hours. This suggest that the interplay between body changes and consciously perceived affect is complex and requires further investigation.

## 2.3 Positive spontaneous and voluntary expressions of emotion

As argued by [63], "the pursuit of happiness is one of the most fundamental human motives". Most of our decisions and actions are driven by the affective forecast of

what option will make us happier. Furthermore, expressing positive affect, via smiling and laughing, can strengthen the bond between two individuals [37].

Among the behavioral cues of positive affect, laughter is commonly targeted as the communicative signal of enjoyment per excellence. Spontaneous laughter is often described as if the people experiencing it abandon themselves to the bodily response of such enjoyment [64]. According to previous research, laughter is composed of respiration, vocalization, body movement, and facial action [64,65]. The facial action is mainly that described as a Duchenne display, or a genuine smiling [64]. Laughter is then accompanied by a series of respiration, vocalization, and body movement bursts. These bursts are often referred as a laughter bout. While laughter is reported to have a mode of four pulses, laughter with one or two pulses also exists. One-bout laughter is called exclamation laughter or chuckle [64]. As the onset of laughter often presents a pre-vocal smiling expression [65], we argue that fast and subtle smiles are similar to a first laughter burst that is quickly contained. In this sense, automatically detecting such facial expressions and laughter in a continuous manner requires the same basic principle.

Whereas most of the autonomic changes cannot be controlled voluntarily, facial expressions and other embodied behavior can be used as communication tool. In this case, the wearer is in control of an affective state or message to be transmitted. Previous research suggested that voluntary and spontaneous affective reactions have different characteristics. They often involve different facial muscles; their temporal dynamics are different; and they are even mediated by distinct neural pathways [37].

Perhaps the most commonly studied facial expressions regarding posed and spontaneous differences are smiles. Besides expressing happiness, a smile can also be used to convey kindness to others. Different terms have been used to refer to these smile-types. Posed and deliberate smiles are often used as synonymous, and they are opposite to spontaneous smiles. Several differences among posed and spontaneous smiles have been found. The most sound difference is the activation of the *orbicularis oculi* muscle that was believed to happen during spontaneous smiles only [37]; in the

so-called Duchenne smile. Nevertheless, recent studies have found that this muscle is activated both in posed and spontaneous smiles [33]. Furthermore, posed smiles tend to have a larger amplitude [34, 38, 39]. Besides these spatial differences, spontaneous and posed facial expressions differ substantially in their temporal dynamics [34,38–42]. Whilst these vary, most agree that posed and spontaneous smiles differ in amplitude, rising and decaying speed, and duration. Spontaneous smiles tend to last longer than posed ones [34,41,42]; they have multiple peaks [42]; and they have longer rising, decaying, and peak durations [38,42]. Furthermore, posed smiles have a longer onset and offset speed [38]. According to [37], spontaneous expressions have a fast and smooth onset; with apex coordination, in which muscle contractions in different parts of the face peak at the same time. In posed expressions, the onset tends to be slow and jerky, and the muscle contractions typically do not peak simultaneously.

Although some studies found that prototypical facial expressions for basic emotions are universal, posed or polite facial expressions might differ depending on the cultural background. [66] showed evidence that when posing smiles, Canadians typically show the Duchenne marker, but Gabonese do not. On the other hand, Mainland Chinese were sensible to the Duchenne marker only when judging smiles from French-Canadians. This suggested that the marker is learned through cultural context. On a follow-up study, [67] explored whether children used the Duchenne marker as a visual cue to distinguish between the two types of smiles. According to their results, children between 4 and 17 years old perceive medium Duchenne smiles as more authentic than equally intense medium non-Duchenne smiles. As they grow older, they rely less on the intensity of the smile. In another study, [68] found that Chinese who use eyes as cue to interpret the facial expression of another person are more accurate than those who use the mouth. Furthermore, those who rated themselves as caring about other people tended to be more accurate and sensitive to the Duchenne marker. Even though these articles support the hypothesis that the ability to pose a smile and to distinguish between posed and spontaneous smiles is acquired through socialization, their definition of spontaneous or genuine smiles is based on the Duchenne marker.

Nevertheless, as they discuss themselves, the ability to display this marker can be learned. Thus, a Duchenne smile is not necessarily spontaneous. This is yet another reason not to rely only on the visual Duchenne marker. The ground truth of the spontaneity of the smile should be established by taking care of the experimental design and the contextual information during the data collection. Despite this, many studies draw conclusions on the perceived spontaneity of a smile using the Duchenne marker [69, 70], and sometimes in static pictures only [71].

## 2.4 Quality assessments through positive affective responses

Recognition of these positive cues could support the fostering and assessment of the mental wellbeing and quality of life of a person or a group, specially of those with developmental disabilities that are unable to effectively communicate [1, 2]. They can also be used as motivational reward in training therapies and tutoring systems [72, 73]. They can be an input to create technology that adapts to human behavior and mental states, such as robots [6]; or to assess the effectiveness of products [3] and media [5, 74]. They can provide better understanding of users and patients during interviews, despite them giving socially desirable answers. This would help both designers of technology and caregivers to adapt their choices to the needs of their users and patients, respectively. Also at a personal level, it would be useful to be aware of our own expressions. Such feedback could be used for social facial expression awareness training, to improve interpersonal communication. This might prove specially useful for persons with autism spectrum disorders (ASD), who have difficulties with social interactions, probably due to difficulties to recognize facial expressions of emotion [75–77]. Finally, it would also be helpful for blind people to perceive their own expressions [58] and the expressions of others. This would create bonding with their interlocutors, and support their social interactions.

Furthermore, the use of technology to automatically measure observable and un-

observable phenomena would increase our temporal resolution to describe and understand embodied cognition. Moreover, being able to quantify these phenomena expands human perception and awareness that could help to better design other agents, products, services, and ultimately ourselves.

# Chapter 3

# Measurement of affective responses

As discussed in previous chapters, quantification of affective responses is useful for different applications. or some of these applications, identifying the presence or absence of affective cues and counting them is enough. Then the interpretation is left for a field expert. However, these perceivable cues are prone to biases and manipulations from the person displaying or reporting the emotion. Therefore, methods to overcome such biases must be considered. This is a challenging task, specially establishing a ground truth. Either with discrete or continuous labels, it is universally accepted that emotion can be either measured through embodied cues or self-report.

Usually embodied cues are labeled by human coders under the assumption that they can have a more objective perception of the facial expression than the persons experiencing the emotion themselves. The person displaying the emotion might be biased to describe what they remember more than what they see. Indeed, one of the biggest challenges of self-reported measures is the subjectivity of emotion itself, and the degree of awareness with which a person can report their emotions.

Sensing with wearable devices can arguably help to overcome those challenges. Sensing technology can measure embodied affective cues more reliably than human coding, and with higher temporal resolution than self-report. However, these technologies require a ground truth label. Such label is, still, assigned by a human coder. When the task is only to identify and count visible behavior, this seems to be the

most appropriate ground truth label. However, when biases are in play, automatic identification must make other judgments beside the perceptual ones. Therefore, an inferential ground truth is required. In this case, the challenge lies in making a correct inference of the affective state of a person.

In this chapter, several methods to measure affective responses are introduced. First, several self-report scales are described. Next, the identification procedure and standards followed by human coders are detailed. Afterwards, computer-aided methods for facial expression identification are mentioned. Then, the challenges of identifying spontaneous expressions automatically is described in more depth. Among those, the ground truth challenge is of special importance.

## 3.1 Self-report

The most straightforward method to investigate how a person is feeling is to ask. Many tools have been developed to reliably self-report affective states. These vary in their definitions of how an emotion should be represented, and on the underlying measurement principle. Dimensional representations include the Affect Grid [13], the Self-Assessment Manikin (SAM) [14], and the Affective Slider [15]. Explicit emotion label reports include the Positive and Negative Affect Scale (PANAS) [78]. Implicit emotion self-reports capitalize on the fact that people tend to make judgments based on their affective state even when they are not aware about their affective state itself. Among the implicit self-report measures we find the Implicit Positive and Negative Affect Test (IPANAT) [79], and the Inkblot test for attitudes [80].

Other studies have shown some evidence that measuring implicit affect adds information to explicit self-report [16]. Nevertheless, self-report has several drawbacks. First, these require time from the user to fill them in. Second, it is unpractical to fill them multiple times to assess experience continuously. This would interrupt the experience itself. Third, the temporal resolution in which we can sample user experience with self-reported tools is limited. This might be because humans are not aware

of their emotional state at every instant. Previous studies have shown that awareness rates of facial expressions are about 60%. Awareness depends on intensity and (or) duration of the facial expression [18]. Scherer (2005) pointed out that, given the complexity of emotion, there are three modes of representing the changes in emotion components: unconscious, conscious, and verbal. The zone of valid self-report is the overlap among these three components. Therefore, there are physiological symptoms, motor expressions, action tendencies, and cognitive appraisals that cannot be fully assessed through self-report [19].

## 3.2 Human coding of affective cues

Facial expression is a basic method that humans have developed to communicate their emotions. Therefore, they are among the most used behavioral cues for emotion recognition.

There are two methodological approaches to study non-verbal behavior. Measuring judgments about one or another message, and measuring the sign vehicles that convey the message [81]. Both approaches involve observers, but observers code the behavior using different criteria. In message judgments, inferences underlying the behavior are made. On the other hand, when measuring the sign vehicles that convey a message, target behaviors are described and counted. These behaviors include counting how many times a muscle moves, or registering the duration of the movement. Even though the labeling task is often simple for a human, it is tedious and time consuming. Moreover, it is usually desired that the human observers describing the behaviors act reliably like a machine would do. However, it is difficult for two human coders to completely agree on the tagged behavior. Even if the judgment is only a perceptual one, there is some error associated with it. This error is caused by perceptual limitations.

Since facial expressions are visually perceivable cues, they can be tagged manually by expert coders. If the coding system is only measuring the sign vehicles that convey

a message, the generated labels are a "perceptual ground truth". When some message is inferred from those cues, the associated labels can be called as "inferential ground truth". As the level of the inferences made increases, agreement among human coders decreases. In the case of inferential ground truths, the quality of the label depends on the ability of the human coder to correctly infer the message or the intention conveyed by a behavior (figure 3.1).

Most studies rely on facial expressions based on the Facial Action Coding System (FACS) [82] to establish a methodology to perform such coding. The primary goal of the FACS is to have a comprehensive reference system which includes all possible visually distinguishable facial movements. The FACS Action Units (AU) were developed to determine the number of muscles which can fire independently, and whether each independent muscular action results in a distinguishable facial appearance. Since video-coding according to the FACS relies on descriptions of behavior units, it is considered as a perceptual ground truth. However, it assumes that the human observer is trained to reliably recognize the specific AUs. Hence, the inference level is higher than just coding facial movement alone.

Finally, it is important to notice that accuracy and error in the context of human coding refers to disagreement. Different coders might disagree on the occurrence of labeled behavior. Therefore, a disagreement between human coders, or between human coders and the machine is used to calculate the accuracy of a rater.

In this thesis, different measures of agreement are used. The Kappa statistic is used to report inter-rater agreement, and accuracy, precision, and recall to measure agreement between the human coder, and the developed machine learning algorithms.

The Kappa statistic (K) is a quantitative measure of the magnitude of agreement between observers [83]. A Kappa of 1 indicates perfect agreement, whereas a kappa of 0 indicates agreement equivalent to chance. The calculation is based on the difference between how much agreement is actually observed, and the agreement that would be expected by chance alone (formula 3.1).

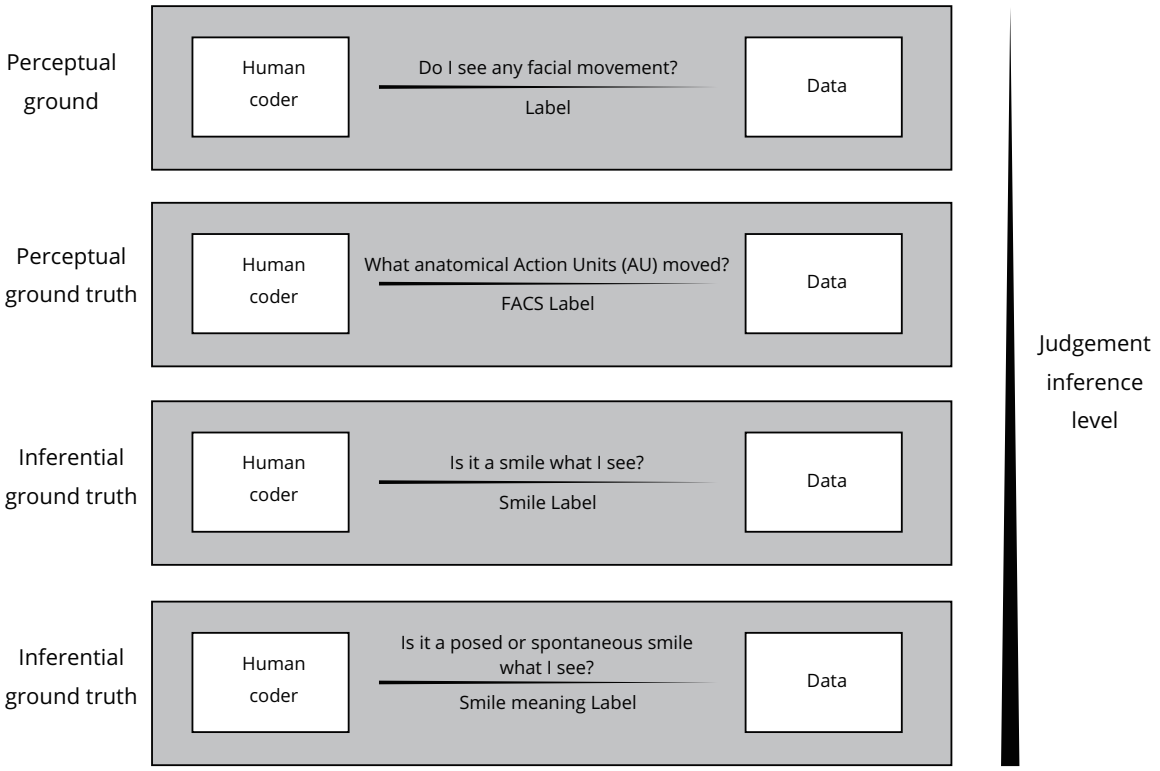



Figure 3.1: *Perceptual and Inferential Ground Truths.* Automatic identification of a behavior depends on a pre-assigned ground truth label. Label can be decided based on human perceptual judgments or inferential judgments. A human coder usually decides the labels, under the assumption that humans are the best at the task at hand.

$$K = \frac{P(A) - P(E)}{1 - P(E)} \tag{3.1}$$

Paradoxically, the Kappa measure is not reliable for rare findings. When the observed behavior is not common, low rates of overall agreement are expected due to the penalization for expected agreement by chance [84].

## 3.3 Computer-aided coding

### 3.3.1 Computer vision-based methods

In the last years, the scientific community has addressed the facial expression recognition challenge using mainly Computer Vision (CV) [28]. Several methods have been developed to detect facial expressions in static images and videos. These methods include tracking of geometric features, often called facial feature points or landmarks; and appearance-based methods using features such as Gabor filters [85–87]. In a review of ten CV-based machine learning studies to recognize emotion, Janssen and colleagues [20] found that most studies rely on facial expressions based on the Facial Action Coding System (FACS) [82]. These studies achieved between 72 and 98% of accuracy. However, the number of features extracted, emotion states detected, and number of subjects participating in the studies varied considerably.

The main advantages of camera-based detection are its high spatial resolution sensitivity; and the unobtrusiveness that can be achieved with the recordings. Despite these advantages, and although the achieved accuracy is relatively high, CV-based facial expression recognition in an uncontrolled environment is still a challenging task. Good recording conditions and having a canonical view of the face are often required [20, 88, 89]. This means that the face should be uncovered and that the person is not allowed to face down or to the sides, which is unnatural. Furthermore, these methods are not robust against occlusion and poor lighting conditions [90]. Recently, some researchers have suggested the use of near-infrared (NIR) video sequences for facial expression recognition to overcome lighting variance [91, 92]. However, occlusion, movement, and point of view challenges remain.

### 3.3.2 Electromyography-based methods

An alternative to camera-based recognition is Surface Electromyography (EMG)-based recognition. EMG has been used extensively in basic research of different fields to quantify facial movement activation [76, 93–95]. Previously, several researchers at-

tempted to use EMG for facial expression detection. In a review of ten EMG-based facial expression detection algorithms, [96] found out a classification performance ranging from 56 till 100%. These studies varied in the number of expressions detected, the number of EMG channels used, the features used, and the classifiers applied. Similar to CV-based methods, comparison among studies is difficult due to the different methodologies pursued. In their exploration, [96] reported a comparison of multiple time-domain EMG features, of which the best performance of 87.1% accuracy was achieved with the Maximum Peak Val ue of the EMG signal.

An advantage of EMG signal is that current EMG technology allows for wireless and compact detection. Therefore, surface EMG could be integrated in a wearable capable of detecting such facial expressions. In [29], an eyebrow emotional expression recognition using surface EMG was proposed. The authors designed a headband embedded with a 10-channel digital EMG. They discriminated between sadness, anger, surprise, disgust, and fear with average accuracy of 96.12%. Another wearable designed to detect facial expressions with EMG was proposed by [30]. In this work, their results show an average precision of 98% for posed smile detection and 96% for posed frowns.

Despite the potential of EMG-based detection, another possible disadvantage is that the EMG magnitude might change over time as a function of fatigue [95]. On the other hand, facial EMG measurement techniques have the advantage of providing an instantaneous, fine-grained muscle activity detection. They are capable of detecting muscle contractions that are too fast or too small to be visually perceived [3, 95, 97].

### 3.3.3 Spontaneous facial expression detection

Most of the automatic detection work previously done is about posed expressions, as opposed to spontaneous expressions. Posed facial expressions are facial expressions collected in a controlled environment when a subject is asked to deliberately produce them. Spontaneous facial expressions are those displayed by freely behaving individuals. These have different characteristics than posed expressions. They often involve

different facial muscles; their temporal dynamics are different; and they are mediated by distinct neural pathways [37]. In recent years, the interest has been shifting to spontaneous facial expressions, as they are more ecologically valid. This is reflected by the recent publication of databases of spontaneous facial expressions [4,88,98–101].

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

## 3.4 Challenges in automatic affective identification

Previous research has shown the feasibility of identifying affective responses automatically using technology. However, many challenges remain. Among those, specially important are the elicitation of ground truth affective states for calibration; the difference between measurements in the laboratory, and the so-called measurements in the wild; and individual differences in affective responses. Ideally, an automatic identification system should be easy to set, and equally accurate during laboratory and in-the-wild measurements. For this purpose, creating a model of affective responses that can generalize to multiple users and situations is of utmost importance. However, this is not an easy task. The first step is to collect data that accurately describes the ground truth. Then, this data can be used to build a model, that can be re-tested with multiple users and situations.

## 3.5 The ground truth challenge

As mentioned before, there are several methods to establish the ground truth on what is the feeling experienced, that matches the embodied affective response of the person. Self-report might be biased. As facial expressions can be fabricated voluntary, assessing perceptual judgments and counting facial expressions only is not enough.

These are also subject to biases as well. Furthermore, in inferential judgments of the meaning of a facial expression, third person video coding judgments might be inaccurate. Therefore, a good experimental design when collecting the calibration data is of utmost importance to ensure that the measured affective responses correspond to one or another affective category.

# Chapter 4

# Smile detection from Facial Electromyography

## 4.1 A wearable device for fast and subtle spontaneous smile recognition

Most of the work previously done in automatic identification of facial expressions is about posed expressions. In recent years, the interest has been shifting to spontaneous facial expressions [74, 88, 98–101]. Due to their different characteristics, spontaneous expressions are more challenging to elicit and detect. These challenges include the detection of fast and subtle spontaneous expressions. Examples of spontaneous, fast, and subtle facial expressions are micro-expressions.

Micro-expressions are brief, subtle, facial expressions that are leaked despite efforts to either deliberately or unconsciously conceal an emotion [31]. They can be considered as spontaneous, because they happen against the will of the person showing them, and better reflect the experienced affect.

*This section is based on:*
Perusquía-Hernández, M., Hirokawa, M., Suzuki, K. A wearable device for fast and subtle spontaneous smile recognition. IEEE Transactions on Affective Computing. Vol. 8, no. 4, pp. 522-533. DOI: 10.1109/TAFFC.2017.2755040

Micro-expressions last only a fraction of a second. Because of their short duration, they are usually neither noticeable to an untrained eye nor to the people disclosing the expression. Besides the concealment, the main element in the definition is the duration. However, there seems to be a lack of consensus in their precise duration range [102–104]. Initially, the boundary between micro and macro-expressions was described as half a second by Ekman and Friesen [31], emphasizing that they are difficult to perceive for the untrained eye.

Previously micro-expressions have been detected mainly through computer vision methods [86, 89, 105–109]. Depending on the number of expressions identified, the results range from 54% accuracy in a leave-one-subject-out validation in [89] to 92% AUC in [109]. On the other hand, other sensors besides video cameras seem to be more robust in different contexts. EMG-based systems have the potential to detect such facial expressions because of their good temporal resolution. High sampling rates make it a promising tool to detect micro-expressions. Moreover, it is robust against head rotations and occlusion; and in a wearable, it could provide independence of movement. Despite this, to the best of our knowledge, micro-expression detection with EMG has not yet been explored. Thus, it remains a question whether these expressions can be identified with the same methods used to identify macro expressions from EMG.

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

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

### 4.1.1 Wearable device

The present work follows the design guidelines for facial expression detection using surface EMG provided by previous research [30], and extensively tested in various settings [110–112]. Figure 4.1 shows the proposed arrangement. It uses four EMG channels placed on the sides of the face, on top of the *temporalis* and the *zygomaticus major* muscles [30]. Since distal EMG is measured, we do not identify the activity of each facial muscle, but a combination of their activities. Therefore, special signal processing is required. The advantage of using distal EMG is that the electrodes do not obtrude the muscle movement. Hence, the facial expressions of interest are not altered by wearing the device. Furthermore, its wearable nature allows for free movement, and good detection in spite of occlusion [112]. However, previous work applied such smile detection only to posed and/or macro-smiles. Micro-smile detection is more challenging due to the temporal and magnitude characteristics of these facial expressions. They are so fast and subtle that even humans experience difficulties perceiving them. As discussed in the following subsections, we propose to improve the design of the wearable device; and to adapt the signal processing for the detection of micro-smiles. The proposed wearable approach is promising for real-time tracking of multiple people's micro-expressions. Usually CV methods are computationally expensive, and more often than not, are limited to the tracking of one person at a time.

### 4.1.2 Garment design

The current system prototype consists of four surface EMG channels connected to a wireless transmitter. The position of the electrodes is on the sides of the face, on top of the temporalis and the zygomaticus major. Each channel consists of two active electrodes bonded together in a 20 by 10 mm box. This box is inserted in a placeholder, which in turn, is attached to a circlet with a bolt and screw. The purpose of the circlet is to keep the four channels in place (See Figure 4.1). This is done by applying pressure on both sides of the face. The four placeholders can rotate slightly

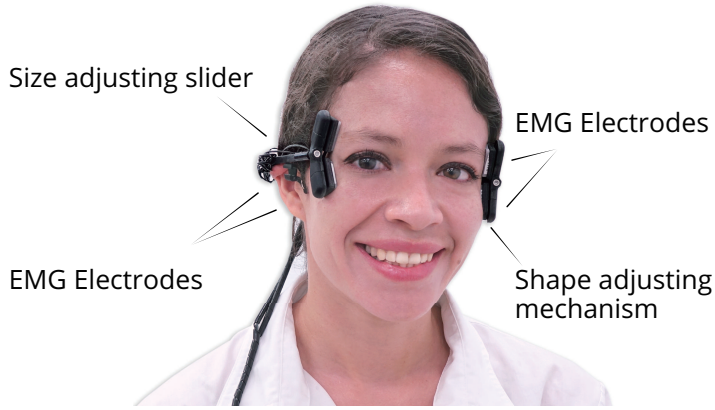



Figure 4.1: The 4-channel EMG wearable used to record the micro-expressions.

inwards, to adjust to the shape of the head of the wearer. Furthermore, the size of the circlet can be adjusted by screwing the attachment between the placeholders and the circlet.

### 4.1.3 Signal processing

The surface EMG is recorded at 1 kHz sampling rate using a four channel Biolog DL-4000 system. The data from all four channels is band-pass filtered from 5 to 350 Hz. Second, it is notch filtered at 50 Hz and its harmonics up to 350 Hz. Next, the signals are decomposed in their Independent Components (IC), using Independent Component Analysis (ICA). The ICA allows to separate the distal EMG from different source muscles. Then the absolute value of the components is considered, and its Root-Mean Square (RMS) value is calculated over overlapping windows of 100 ms, sliding one sample at a time. This was done to increase the temporal resolution of the algorithm, hence optimizing it for micro-smile detection. The aforementioned pre-processing was performed on all EMG data, for each participant. The resulting data is considered as input features to train a Neural Network (NN) with one hidden layer of four Sigmoid neurons (Figure 4.2). Due to anatomical differences in muscle size and Body Mass, EMG is highly variable between subjects. Hence, within subject data was used to train the NN. Furthermore, given the limited availability of micro-smile samples, the no-expression data was under sampled to match the number of samples

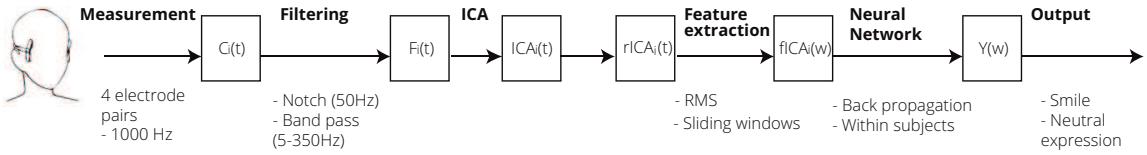



Figure 4.2: Signal processing for micro-smile detection steps per participant.

of the expression data [113,114]. No-expression data was taken randomly from all the data. To validate this model, cross-validation with 70% train, 15% validation, and 15% test data is used. The neural network aims to compare micro-expressions with no-expression display. Micro-smiles are compared to no-expressions, as the electrode positions are optimized for positive expressions.

### 4.1.4 Experiment 1

The main purpose of this experiment is to assess the possibility of detecting fast and subtle smiles with distal EMG. To the best of our knowledge, no online database is available, which includes unobtrusive, distal EMG of micro-expressions, recorded in the proposed arrangement. Therefore, an experiment was designed for data collection. To elicit micro-expressions, a methodology similar to the one described in [88] was used. Furthermore, the video stimuli were mainly Ads that could be assessed with the proposed method. The experiment was within-subjects, where all participants watched all stimuli in a counterbalanced order.

#### Stimuli selection

Three Ad videos were selected from stimulus used in previous research [74]. Namely, “The force” (Video TF, 62 s), “House sitting” (Video HS, 30 s), “Parisian Love” (Video PL, 53 s). According to McDuff et al., using thiese stimuli they could collect more than 10 000 frames of smiles, hence we expected to get similar results. All videos were presented at 30 frames per second with 720x480 pixel resolution.

### Pilot

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

### Participants

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

### Procedure

Participants were provided with a general description of the test. The description stated that the purpose of the test was to rate several videos, while recording their facial EMG. Second, they provided their informed consent. Next, they were given the right to quit at any time. The task consisted in watching three videos in counterbalanced order. Before each video, they were asked to watch 30 seconds of black screen, as a baseline for the measurements. They were instructed to "keep a neutral face while watching the videos". Finally, the experimental setup is shown in Figure 4.3.

### Measurements

During the task, surface EMG and the participant's face were recorded simultaneously. The camera was a Canon Ivis HFg10 with HD resolution at 30 frames per second (fps) for the first five participants. For the last seven participants, the camera was changed to a Sony Cyber-shot DSC-RX10 II with 1920 x 1020 resolution at 120 fps. The purpose of this was to increase the number of frames in which the

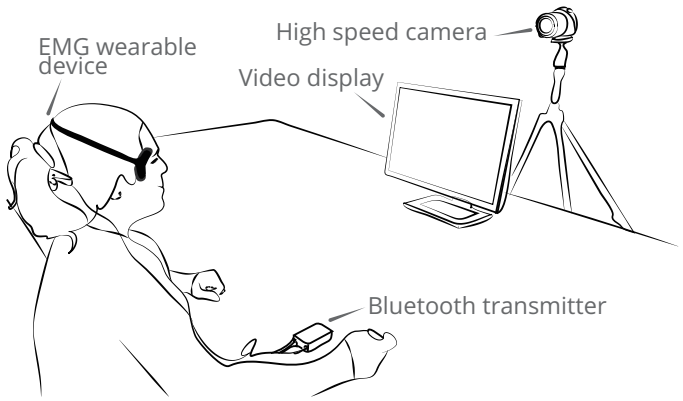



Figure 4.3: *Experiment setup for micro-smile collection.* Participants wore the EMG device while sitting in front of a LCD screen where the stimuli videos were shown. A camera was placed in front of them to record their facial expressions. This video was used as ground truth.

micro-smile would be shown, and therefore facilitate the job of the human coder. As mentioned before, a micro-expression can be argued to last as little as 1/25 of a second. Even with a normal camera recording at 30 fps, there would be at least one recorded frame of each micro-smile lasting 1/25 seconds. The number of frames is increased to four in the case of a 120 fps high-speed camera.

### Results experiment 1

**Video coding.** All the recordings of the participants' face were coded frame-by-frame for facial expressions by three experienced coders. High-frame rate videos were slowed down to facilitate the task. The labeling included coding for the onset, offset, and apex frames of the facial expression; the Facial Action Unit System's (FACS) Action Units that were present in the expression; and whether it was considered a smile or not, and laughter or not. The labeled AU were AU1, AU02, AU04, AU05, AU6, AU09, AU10, AU12, AU14, AU15, AU17, AU18, AU25, AU26, AU28, and AU38. Smiles were often a display of AU6 and/or AU12. However, the smile label was not assigned every time these AU occurred [74]. A facial expression was considered as each annotation of the appearance of an AU change from the face on resting state or a change from a different AU. The duration of the expression was calculated as the

Table 4.1: *Number of expressions elicited per video.* Video 1 lasted for 62 s, Video 2 for 30 s, Video 3 for 53 s, and Video 4 for 129 s. Micro-expressions are those lasting for less than 0.5 s. Other expressions refer to facial expressions that were not labeled as smiles.

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

difference between the coded onset and the apex. All facial movements considered as swallowing, coughing, or sneezing were excluded.

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

Table 4.1 shows the number of expressions that were elicited by each video. From the elicited expressions, 238 were smiles; 177 were expressions faster than one second; 67 of the smiles were faster than one second; and 95 expressions were faster than 1/2 second, from which 32 were smiles. Expressions lasting less or equal than half a second were considered micro-expressions.

The Cohen's Kappa Coefficient was used as a measure of inter-rater agreement [83]. For this paper, only the information of the duration of the expression plus the assessment of whether the expression was a smile or not were used. Therefore, the Kappa Coefficient was calculated on the frame-by-frame human coding on whether the participant was smiling or not. According to this, the Cohen's Kappa Coefficient was 0.4068 ($p<0.01$) between rater 1 and 2; 0.33 ($p<0.01$) between rater 1 and 3; and 0.54 ($p<0.01$) between rater 2 and 3. Furthermore, the Fleiss' Kappa coefficient [115] among all raters was 0.41 ($p<0.01$).

**EMG signal processing.** Figure 4.4 shows the EMG processing steps for participant seven, video four. Dark dotted squares indicate the smile labels given by the human coder. Four examples of the facial expressions are also shown. In here, only

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

Sixteen out of 23 participants displayed micro-expressions, and 10 of them displayed micro-smiles. Only their data was used for further analysis. Table 4.2 shows the precision and recall achieved for micro-smile detection for each participant, for the neural network results between micro-smiles and no-expression. Table 4.3 shows similar results but for classification of smiles lasting at most one second. Interestingly, the achieved performance seems to decrease as longer smiles are included in the data set.

### 4.1.5 Experiment 1-b

EMG-based facial expression detection has the potential of being used to annotate the interaction between Ads or other stimuli, and human-human interactions. Previously, [74] showed that smiles detected using CV over the internet can be used to predict online media effectiveness. However, this study was limited to macro-smiles of a single viewer. As shown in previous sections, EMG has critical advantages in detecting micro-smiles in real time, and where occlusion is likely to occur. In this section, we show an exploratory study on the performance of simultaneous detection of multiple people‘s fast and subtle smile detection using our proposed method.

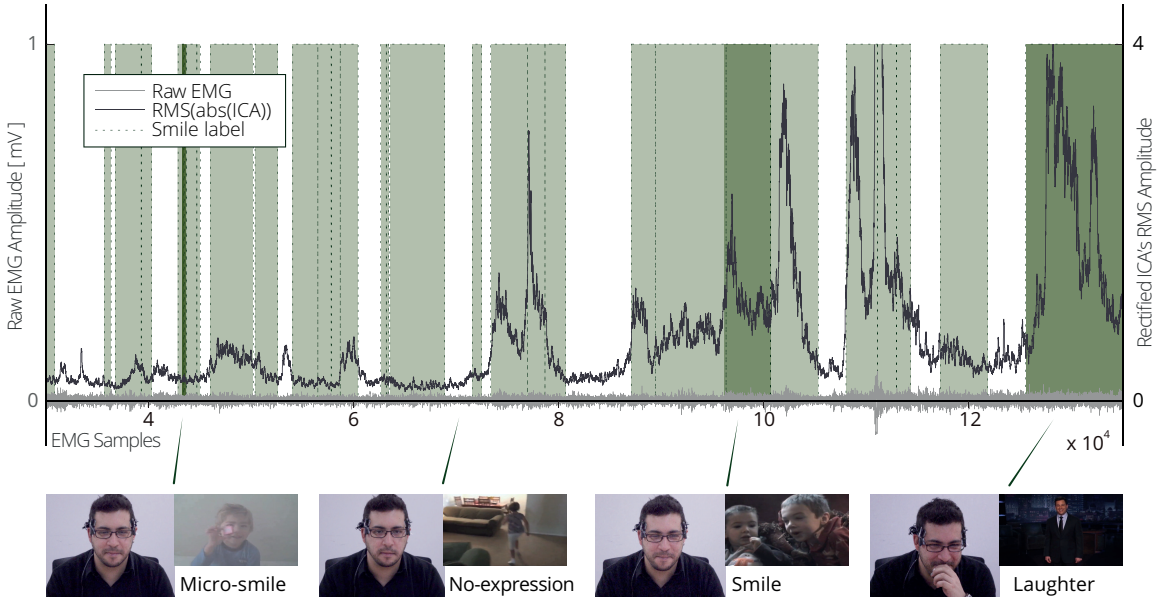



Figure 4.4: *Processing stages for smile detection.* The dark dotted squares indicate the smile labels given by a human coder. An example frame shot of the coded expression is given below the plot. The stimuli video frame seen at that moment is also shown. The gray signal is one of the four raw EMG channels, and its magnitude is shown in the left Y axis. The black signal corresponds to the RMS of one of the four rectified ICA components. Its magnitude is shown in the right Y axis. The latter was used as input for the Neural Network. Axis X represents the EMG sample number, as synchronized with the video.

Table 4.2: *Micro-smile and neutral faces classification results.* Micro-smiles were lasting less than 0.5 s. RMS was used as feature. The number of smiles represent the number of smiles observed. The number of smile EMG samples represents the total amount of EMG samples taken at 1 kHz from onset till offset of all reported smiles. Classification results of micro-smiles. RMS was used as feature.

| Subject | Micro-smile detection | | | Number of smile EMG samples | Number of smiles | Age | Gender |
|---|---|---|---|---|---|---|---|
| | Precision | Recall | Accuracy | | | | |
| 1 | NA | NA | NA | 0 | 0 | 25 | Male |
| 2 | 100% | 100% | 100% | 434 | 1 | 26 | Female |
| 3 | NA | NA | NA | 0 | 0 | 27 | Male |
| 4 | NA | NA | NA | 0 | 0 | 24 | Male |
| 5 | NA | NA | NA | 0 | 0 | 35 | Male |
| 6 | 100% | 100% | 100% | 418 | 1 | 26 | Male |
| 7 | 98.1% | 98.4% | 98.2% | 1855 | 7 | 34 | Male |
| 8 | NA | NA | NA | 0 | 0 | 34 | Male |
| 9 | NA | NA | NA | 0 | 0 | 29 | Female |
| 10 | 99.8% | 99.9% | 99.8% | 2318 | 10 | 28 | Male |
| 11 | NA | NA | NA | 0 | 0 | 28 | Female |
| 12 | NA | NA | NA | 0 | 0 | 28 | Male |
| 13 | 79.3% | 82.4% | 81.20% | 1160 | 3 | 27 | Female |
| 14 | 100% | 100% | 100% | 243 | 1 | 24 | Male |
| 15 | 100% | 100% | 100% | 434 | 2 | 26 | Male |
| 16 | 94.6% | 96.8% | 95.8% | 1605 | 4 | 23 | Male |
| 17 | NA | NA | NA | 0 | 0 | 31 | Female |
| 18 | NA | NA | NA | 0 | 0 | 24 | Female |
| 19 | 100% | 100% | 100% | 401 | 1 | 23 | Male |
| 20 | 100% | 100% | 100% | 286 | 2 | 22 | Male |
| 21 | NA | NA | NA | 0 | 0 | 24 | Male |
| 22 | NA | NA | NA | 0 | 0 | 25 | Male |
| 23 | NA | NA | NA | 0 | 0 | 25 | Female |
| Total | | | | 9113 | 32 | | |

Table 4.3: *Smiles lasting less than a second and neutral faces classification results.* RMS was used as feature. The number of smiles represent the number of smiles observed. The number of smile EMG samples represents the total amount of EMG samples taken at 1 kHz from onset till offset of all reported smiles. Classification results of micro-smiles RMS was used as feature.

| Participant | Micro-smile detection | | | Number of smile EMG samples | Number of smiles |
|---|---|---|---|---|---|
| | Precision | Recall | Accuracy | | |
| 1 | 94.9% | 96.9% | 95.9% | 801 | 1 |
| 2 | 96.3% | 99.6% | 98.00% | 1167 | 2 |
| 3 | NA | NA | NA | 0 | 0 |
| 4 | NA | NA | NA | 0 | 0 |
| 5 | NA | NA | NA | 0 | 0 |
| 6 | 79.8% | 88.1% | 84.5% | 3705 | 5 |
| 7 | 89.5% | 94.9% | 92.3% | 8356 | 17 |
| 8 | 98.5% | 93.5% | 95.8% | 926 | 1 |
| 9 | NA | NA | NA | 0 | 0 |
| 10 | 82.8% | 91.7% | 87.6% | 5041 | 14 |
| 11 | 100% | 100% | 100% | 2045 | 3 |
| 12 | NA | NA | NA | 0 | 0 |
| 13 | 99.9% | 99.5% | 99.7% | 3940 | 4 |
| 14 | 100% | 100% | 100% | 1128 | 2 |
| 15 | 100% | 100% | 100% | 393 | 2 |
| 16 | 64.7% | 80.3% | 74.4% | 5895 | 11 |
| 17 | 83.1% | 81.1% | 81.9% | 651 | 1 |
| 18 | NA | NA | NA | 0 | 0 |
| 19 | 100% | 99.9% | 99.9% | 2754 | 2 |
| 20 | 100% | 100% | 100% | 286 | 2 |
| 21 | NA | NA | NA | 0 | 0 |
| 22 | NA | NA | NA | 0 | 0 |
| 23 | NA | NA | NA | 0 | 0 |
| Total | | | | 37088 | 67 |

### Experiment Design

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

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

### Participants

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

### Stimuli

For the micro-smile elicitation block, the same videos from the previous experiment were used. Additionally, three new Ad videos were selected, plus a video showing funny and cute kid behavior. The videos are, "Baby expectancy" (Video 5, 29 s), "Fun kiddies" (Video 6, 2 min 18 s), "Brotherhood" (Video 7, 53 s), "Dirt Devil" (Video 8, 1 min 28 s). All videos were presented to the participants at 30 frames per second with 720x480 pixel resolution.

### Measurements

During the task, surface EMG and the face of both participants were recorded simultaneously. The camera used was a Sony Cyber-shot DSC-RX10 II with 1920 x 1020 resolution at 120 fps. For device-device-stimuli synchronization purposes, a hardware trigger was designed. This consisted of a micro-controller interfacing via

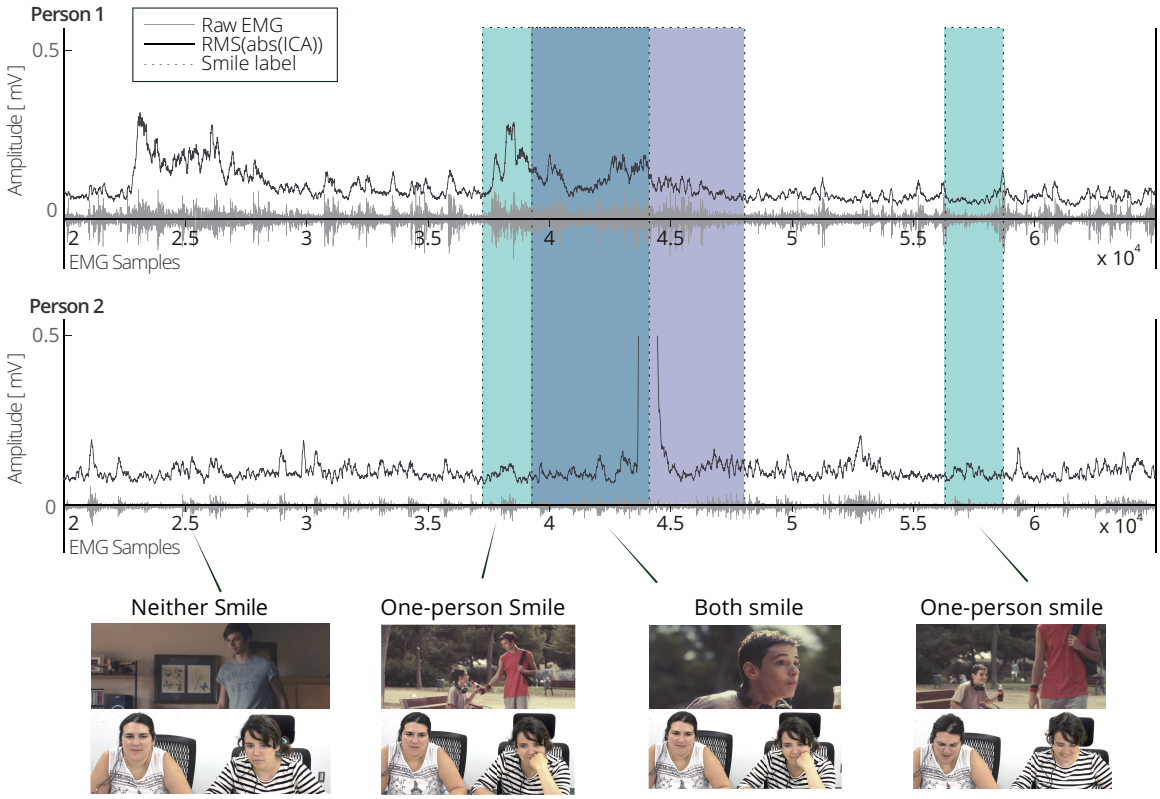



Figure 4.5: *Processing stages for smile detection in a multi-user setup.* The wearable can also be used to measure multiple users at the same time, a hardware trigger was used to synchronize multiple devices and the stimuli presentation. With this method, it can be observed when people smile simultaneously or alone.

serial port with a laptop used to present the stimuli and send the triggers to the EMG acquisition device.

### Results

All the recordings of the participants' face were coded frame-by-frame for facial expressions by one experienced coder, similarly as in the previous study. From the elicited expressions, 166 were smiles, 76 were expressions faster than half a second, and 24 were smiles faster than half a second (Table 4.4). In this study, six out of eight participants tended to rotate their heads to face the other participant, to cover their faces depending on the situation and to talk quite often. Figure 4.5 shows the results of human coding for totally and partially covering the face, looking towards each

Table 4.4: *Number of expressions elicited per video in experiment 1-b.* Video 1 lasted for 62 s, Video 2 for 30 s, Video 3 for 53 s, Video 4 for 129 s, Video 5 for 30 s, Video 6 for 158 s, Video 7 for 57 s, and Video 8 for 88 s. Micro-expressions are those lasting for less than 0.5 s. Other expressions refer to facial expressions that were not labeled as smiles.

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

Table 4.5: *Micro-smiles lasting less than 0.5 s and neutral faces classification results for experiment 1-b.*

| Participant | Spontaneous smile detection | | | Number samples | Number expressions | Age | Gender |
|---|---|---|---|---|---|---|---|
| | Precision | Recall | Accuracy | | | | |
| 1 | NA | NA | NA | 0 | 0 | 29 | Male |
| 2 | NA | NA | NA | 0 | 0 | 31 | Male |
| 3 | NA | NA | NA | 0 | 0 | 30 | Female |
| 4 | 96.8% | 97.5% | 97.2% | 493 | 1 | 29 | Female |
| 5 | 100% | 100% | 100% | 1120 | 4 | 30 | Female |
| 6 | NA | NA | NA | 0 | 0 | 30 | Male |
| 7 | 88.8% | 91.3% | 90.2% | 3686 | 17 | 25 | Female |
| 8 | 100% | 100% | 100% | 886 | 2 | 30 | Male |
| Total | | | | 6185 | 24 | | |

other, and looking away to the camera. Smile labels are also shown in the figure. Tables 4.5 and 4.6 summarizes the results of the classification with the proposed method for micro-smiles and for longer-lasting smiles. Here, 2 s or less was chosen arbitrarily to exemplify how performance varies with the duration. Longer smiles and laughter present more variation in the EMG due to muscle fatigue or multiple EMG bursts, respectively. Therefore, they are more difficult to identify as a unit. Future work should explore possibilities to address these challenges.

Table 4.6: *Micro-smiles lasting less than two seconds and neutral faces classification results for experiment 1-b.*

| Participant | Spontaneous smile detection | | | Number samples | Number expressions |
|---|---|---|---|---|---|
| | Precision | Recall | Accuracy | | |
| 1 | NA | NA | NA | 0 | 0 |
| 2 | 97.7% | 99.0% | 98.4% | 5529 | 4 |
| 3 | 96.1% | 95.1% | 95.6% | 2619 | 2 |
| 4 | 87.4% | 93.3% | 90.6% | 3353 | 3 |
| 5 | 86.3% | 87.2% | 86.8% | 58050 | 28 |
| 6 | NA | NA | NA | 0 | 0 |
| 7 | 79.3% | 84.8% | 82.6% | 19074 | 39 |
| 8 | 85.1% | 92.3% | 89.0% | 2521 | 7 |
| Total | | | | 91146 | 83 |

**Discussion and conclusions**

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

The achieved accuracy to identify micro-smiles was very good. This high accuracy cannot be explained due to overfitting, as we made sure to apply proper cross-validation. Nevertheless, an important challenge is the unbalanced nature of the data. Micro-smiles are very rare compared to the no-expression class. To address this issue, we under sampled the non-expression data. The results shown in Table 2a are reporting the accuracy for this case. Interestingly, the same algorithm yielded less accurate results if the inclusion criteria was extended to fast and subtle smiles lasting less than a second (Table 2b). In the latter case, accuracy ranged from 74 till 100% depending on the participant. This could be explained by the duration of the expressions themselves. As the duration time is shorter, there are less variations in the EMG intensity as can be observed in longer expressions. In longer smiles, the relationship between detected electrical output at an EMG site, and the mechanical force exerted by a muscle may change over time as a function of fatigue [95]. This variability would make it more difficult for the machine learning algorithm to define a clear boundary between samples of different expressions. Furthermore, previous work has suggested that muscle contractions in spontaneous expressions peak simultane-

ously [37], which could contribute to the success of the EMG-based classification. An alternative explanation could be that usually smiles happen on top of other smiles. It was observed that once some participant smiles, the smile tends to last for a long period of time, and after a while, this smile is further extended in reaction to a funnier stimuli event. This could also contribute to the clear magnitude difference between the neutral expression and a micro-smile.

Furthermore, the temporal resolution and portability of this device would allow to provide real-time feedback, if desired. This can be used for quantification applications of positive facial expressions. In spite of using a 100 ms window for processing, the temporal resolution is still 1kHz, because the window is sliding every sample. However, there is still some work to be done before bringing this wearable device to a real-time application setting. Only the results of an offline setup are showed here. For real-time applications, the micro-smile recognition should be ported to an online classification system. Moreover, one of the challenges would be the calibration of the system for each individual. Previous work on CV shows that inter-subject training and validation leads to poor results [89]. Regarding EMG, high inter-subject and inter-session variability is also expected. This is mainly due to anatomical differences in muscle size and position, and in Body Mass Index. In between sessions, one of the main sources of error are the differences in electrode position [30,116], and changes in skin conductance; suggesting that individual calibration is preferred to ensure good performance. However, elicitation of such fast and subtle spontaneous expressions for calibration purposes is a challenging task.

Previously, Yan et al. [117] discussed that the stronger the emotion felt, the more micro-expressions could be elicited. Furthermore, a high stakes situation is often required to elicit micro-expressions. In our case, the videos were short, they were not watched for the first time, and they elicited mild emotional content. Furthermore, the intensity of the facial expressions has also been argued to depend on the elicitation paradigm used. The neutralization paradigm implies that facial expressions are inhibited with strong intent, and therefore not leaked easily. On the other hand,

during our tests, we could observe the two extremes: participants 4 and 9 did not leak any expression, whereas most of them were showing wide smiles. Other participants broke in laughter in several occasions. This resulted in all their smiles lasting longer than half a second. This phenomenon was observed before by [117]. They reported that expressions of happiness tend to last longer than the micro-expression threshold, and that micro-expressions of disgust are much more frequent. In our data, 34.3% of the elicited micro-expressions were smiles, despite the stimuli being rated as positive, and the smiles being 63.5% of the macro-expressions. Even though with the proposed experimental setup, only about 44% of the participants displayed micro-smiles. This is not unexpected. In their experiment Yan et al. also discussed that "As for micro-expressions of happiness, one may feel surprised as to why so few were elicited. Though happiness feelings were easily elicited when watching amusing video episodes, these elicited smiles or laughter bursts are lasting facial expressions and do not fit the criteria of micro-expression. Thus, most of the elicited happy facial expressions were categorized as conventional facial expressions." The phenomena we observed is similar, despite using different stimuli. Furthermore, when the stimuli are strong enough, the displayed affective expression is often long-lasting laughter. It requires quite some effort to conceal laughter, and while some participants are very good in neutralizing their facial expressions, most others are not. Therefore, we argue that a fast and subtle smile is a first laughter burst that is quickly contained, and that the proposed method and device is thus able to detect both.

Despite the number of micro-smiles might seem small, the high sampling frequency of EMG allows to obtain 500 samples (at most) per micro-smile. In other words, we could analyze about 9113 EMG samples of micro-smiles in the single-participant experiment, and 6185 in the pairs-experiment.

Even though this was the first effort to evaluate the detection of fast and subtle spontaneous facial expressions using an EMG-based wearable device, we only tested with smiles. Future work should address the possibility of extending the detection to more facial expressions. Smiles were a good first step because of their potential

in different application domains, specially as a measure of wellbeing; as measure of acceptation of a product or therapy; or as a positive reward in learning applications.

The current wearable device electrode positions were selected based in [30], and are optimized for smile detection. This was done in the aforementioned research by considering a multi-attribute decision making process, facial morphology, and EMG amplitude. Moreover, the device is able to detect other expressions such as posed frowning. Including more electrodes would be helpful to expand the wearable's spatial resolution to detect more facial expressions. In the future, the use of dense and compact grids of electrodes around the face [116] seem a good alternative to achieve so without covering it. Covering the face is an undesired situation as it makes the users aware of their own facial expressions, and limits the movement of the skin [30, 95].

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

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

We presented a wearable device which can provide laughter analysis in human-

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

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

As outlined in previous chapters, only identifying and counting smiles is not enough to assess the affective experience of the person displaying the smile. It is also important to be able to infer some affective meaning from the facial expression. Since smiles can be voluntarily fabricated, an algorithm that can differentiate between posed and spontaneous smiles would be useful.

*This section is based on:*
Perusquía-Hernández, M., Hirokawa, M., Suzuki, K. Spontaneous and posed smile recognition based on spatial and temporal patterns of facial EMG. 7th Affective Computing and Intelligent Interaction.

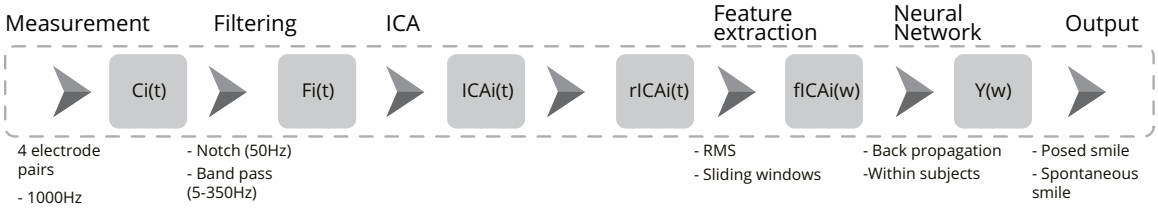
In this study, we propose to use EMG wearable technology to explore the temporal characteristics of spontaneous and posed similes. 16 participants were selected from the previous experiments. These participants were among the ones that showed both spontaneous and posed expressions.

To distinguish between posed and spontaneous smiles, two algorithms were designed. The first one analyzes spatial and magnitude features of facial EMG. In addition, a temporal feature calculation algorithm is added to the second method to assess the differences in temporal dynamics between both types of smiles.

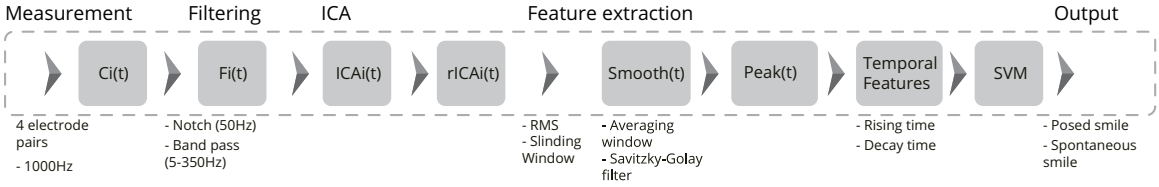
It is important to note that the wearable device used to record the EMG signals consisted of four wireless surface dry-electrode EMG channels. These four electrodes were placed in a circlet that can be comfortably worn without covering the face (Figure 4.1). With this wearable, the electrodes were placed on the side of the face, on top of the *temporalis* and the *zygomaticus major* muscles. As previously described in section 4.1.1. This configuration was proven to be robust for detection of smiles and frowns in various settings [110–112]. Furthermore, this arrangement uses the distal electrode locations on the side of the face in order to capture facial expressions. Hence, the activity of each facial muscle is not identified, but the conducted pattern classification considers facial expressions as a combination of the activity of all related facial muscles. Due to the overlapping of the distal signal from different muscle groups, special biosignal processing is applied.

### 4.2.1 Data collection

1. **Experiment design and procedure.** Participants were informed that the purpose of the test was to rate some videos with a questionnaire and by measuring their facial EMG. To elicit spontaneous smiles, a series of fun video stimuli were showed to the participants in a counterbalanced order. In total, participants watched eight videos. During the first four, they were asked to "keep a neutral face while watching the videos". Therefore, we expected all leaked expressions to be spontaneous. In the second block, no particular instruction




(a) Spatial and magnitude features analysis pipeline.




(b) Temporal feature analysis pipeline.

Figure 4.6: *The signal processing steps to distinguish between posed and spontaneous smiles.*

was provided. Before and after watching the stimuli, participants were asked to pose smiles, frowns, and eye-brow lifts, supposedly with the purpose of verifying the EMG signal. After the experiment, participants were debriefed.

2. **Stimuli.** Three Ad videos known to elicit smiles were selected from previous research [74]: "The force" (Video TF, 62s), "House sitting" (Video HS, 30s), "Parisian Love" (Video PL, 53s). Additionally, an edition of the 2011 Jimmy Kimmel Challenge "I Told My Kids I Ate All Their Halloween Candy" (Video HA, 2min 9s), was included. During these four videos, participants were asked to avoid making any facial expressions. Four extra videos showing fun and cute behavior were watched with no particular instruction: "Baby expectancy" (Video 5, 29s), "Fun kiddies" (Video 6, 2min 18s), "Brotherhood" (Video 7, 53s), "Dirt Devil" (Video 8, 1min 28s). All videos were presented at 30 frames per second with 720x480 pixel resolution.

3. **Participants.** Sixteen voluntary participants took part on the study (average age=26.3 years old, SD=3.24, 6 female). Eight participants were Japanese and the rest from other European and Latin American countries. None of them had experience using the measuring device, and 11 participants had seen at least

one of the videos before.

4. **Measurements.** During the task, surface EMG and the participant's face were recorded simultaneously. The surface EMG was recorded at 1 kHz sampling rate using a four channel Biolog DL-4000 system. The camera used was a Sony Cyber-shot DSC-RX10 II with 1920 x 1020 resolution at 120 fps.

### 4.2.2 Data analysis

#### Video coding

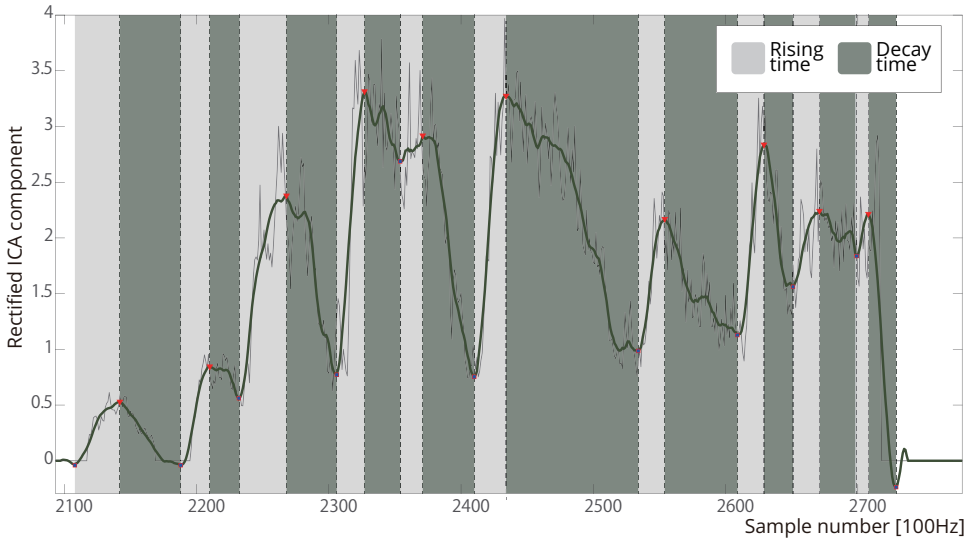
All the recordings of the participants' face while watching the stimuli were coded frame-by-frame for facial expressions by two experienced coders. The labeling included coding for the onset, offset, and apex frames of the facial expression; the Facial Action Unit System's (FACS) Action Units that were present in the expression; and whether it was considered a smile or not, a posed expression or not, and laughter or not. Smiles were often a display of AU6, AU12 and/or AU25. However, the smile label was not assigned every time these AU occurred [74]. All facial movements considered as swallowing, coughing, or sneezing were excluded. For the posed expressions block, the instruction given to the participant was used to label the expression as a smile or not. Furthermore, an experienced coder labeled the data in the same manner as described for the stimuli block to identify the start and the end of the posed facial expressions.

#### Signal processing

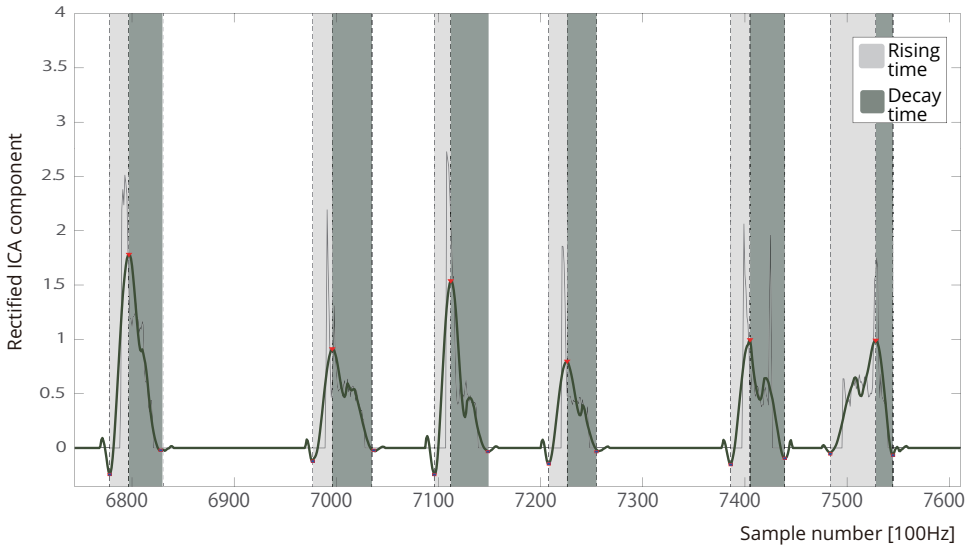
The EMG signals picked up by the electrodes are transmitted to a laptop via Bluetooth, where they are analyzed using Matlab 2014a. Two different detection algorithms are proposed. The first one relies on magnitude and spatial features only, whereas in the second, temporal features of the signal were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

**Spatial and magnitude features analysis pipeline.** The data from all four channels is band-pass filtered from 5 to 350 Hz. Second, it is notch filtered at 50 Hz and its harmonics up to 350 Hz. Afterwards, any linear trends were removed from the signal. Next, Independent Component Analysis (ICA) was applied, to separate the components from different muscles. Then, the absolute value of the resulting components was calculated. Next, the RMS value is calculated over overlapping windows of 100 ms, sliding one sample at a time. The resulting data was labeled according to the human coding, and used as features to train a Neural Network (NN) in cross-validation with 70% train, 15% validation, and 15% test data. The neural network aims to compare spontaneous smiles with posed smiles. Due to the unbalanced nature of the data, the majority class was undersampled to match the samples of the minority class [113, 114]. This process is shown in figure 4.6(a).

**Temporal feature analysis pipeline.** This method followed a similar preprocessing as the previous one (Figure 4.6(b)). The four channels were band-pass filtered (5-350 Hz); notch filtered (50 Hz and harmonics); de-trended; the absolute value of its ICAs was calculated; and the RMS of the signal was calculated using an overlapping 100 ms window sliding every sample. Afterwards, smile data was selected according to the human label, and sliced in individual smiles. Next, each smile data was smoothed using an averaging non-overlapping window of 100 ms, and a Savitzky-Golay Filter with a 5th order polynomial and 41 as frame length. Then, peak detection was performed on the smoothed EMG signal to calculate the rising, and decay times (Figure 4.7). The rising time is defined as the time taken from the first minimum to the first maximum (peak) in the smoothed signal; decay time is defined as the time between the last maximum to the last minimum. Furthermore, the magnitude change during rising and decay; and the rising and decay speed were calculated as features. The resulting feature set was standardized and used to train a Support Vector Machine (SVM) to distinguish between posed and spontaneous smiles. A Gaussian Kernel Function was used. To validate the model, a cross-validation with 70% train, 15%




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




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

Figure 4.7: *The EMG signature of a smile.* A sample of smile data with the estimated envelopes, peaks, and rise and decay sections. No-smile EMG data was masked with zeros for easy visualization.

validation, and 15% test data was used. As with the other method, the data was balanced to match the minority class.

Table 4.7: *EMG spatial and magnitude features-based posed and spontaneous smiles identification.* RMS was used as feature as described in section 4.2.2

| Participant | Spontaneous-smile detection | | |
|---|---|---|---|
| | Precision | Recall | Accuracy |
| 1 | 75.20% | 85.90% | 81.40% |
| 2 | 68.00% | 64.90% | 65.60% |
| 3 | 78.90% | 77.20% | 77.80% |
| 4 | 68.50% | 80.00% | 75.70% |
| 5 | 91.80% | 93.30% | 92.60% |
| 6 | 83.00% | 72.40% | 75.70% |
| 7 | 78.60% | 70.90% | 73.10% |
| 8 | 98.50% | 97.20% | 97.80% |
| 9 | 33.00% | 61.30% | 56.10% |
| 10 | 80.50% | 71.20% | 73.90% |
| 11 | 71.00% | 74.40% | 73.30% |
| 12 | 94.90% | 88.50% | 91.30% |
| 13 | 92.00% | 80.50% | 84.80% |
| 14 | 92.20% | 87.50% | 89.50% |
| 15 | 83.70% | 76.70% | 79.10% |
| 16 | 73.80% | 81.60% | 78.60% |
| Average | 74.19% | 77.80% | 77.75% |
| SD | 15.07% | 9.58% | 10.18% |

### 4.2.3 Results

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

The spatial and magnitude features analysis pipeline was moderately successful in distinguishing spontaneous and posed smiles. The classification results range from 56.10% till 97.80% of accuracy. On the other hand, the temporal feature analysis pipeline accuracy values ranged from 85.23% till 96.43%. A Wilcoxon Signed-Rank test showed that the spatio-temporal features algorithm's accuracy is significantly higher than the one using spatial and magnitude features ($V=9$, $p < 0.01$). This clas-

Table 4.8: *EMG spatio-temporal features-based posed and spontaneous smiles identification.* Spatio-temporal features were used as described in section 4.2.2

| Participant | Spontaneous-smile detection | | |
|---|---|---|---|
| | Precision | Recall | Accuracy |
| 1 | 85.16% | 97.32% | 87.98% |
| 2 | 90.00% | 100.00% | 90.70% |
| 3 | 97.65% | 94.32% | 95.33% |
| 4 | 87.50% | 78.87% | 85.63% |
| 5 | 96.30% | 100.00% | 96.43% |
| 6 | 90.44% | 86.62% | 88.45% |
| 7 | 87.94% | 100.00% | 89.43% |
| 8 | 92.05% | 99.29% | 92.74% |
| 9 | 91.84% | 60.81% | 85.23% |
| 10 | 91.67% | 72.13% | 88.40% |
| 11 | 86.13% | 93.13% | 88.18% |
| 12 | 90.50% | 87.10% | 88.67% |
| 13 | 84.75% | 89.82% | 86.54% |
| 14 | 87.55% | 90.27% | 88.08% |
| 15 | 89.37% | 93.91% | 89.41% |
| 16 | 84.62% | 84.62% | 86.29% |
| Average | 89.44% | 87.73% | 89.11% |
| SD | 3.69% | 10.66% | 3.10% |

sification was made in an average of 350 temporal features per participant extracted from the envelopes of the EMG measured while smiling. From these features, a series of t-tests revealed that spontaneous smile duration differs from posed smile duration ($t(2276)=-11.535$, $p<0.01$). Second, the magnitude both types of smiles is not significantly different ($t(2276)=-0.19837$, $p>0.05$). Third, the rising time ($t(1151)=-7.5336$, $p<0.01$) and decay time ($t(1124)=-8.8359$, $p<0.01$) differ, but the speed of change is not significant. Neither during the rising phase ($t(1130)=0.22068$, $p>0.05$) nor during the decaying phase ($t(1108)=1.6413$, $p>0.05$).

Finally, no significant cultural differences were observed. However, figure 4.8 shows that the spatio-temporal features result for Japanese tend to have less variation than for other nationalities.

### 4.2.4 Discussion and conclusion

EMG has the potential to distinguish between spontaneous and posed smiles in a portable and real-time fashion using a wearable. We provided support for this using both magnitude and temporal features. Our proposed algorithms take advantage of the ICA extraction to estimate different sources of the EMG signal and its magnitude. They also profit from the EMG's high temporal resolution to estimate smile characteristics without consuming excessive computational resources. From these two alternatives, the most successful results were given by considering the temporal resolution of the signal. As supported by previous studies, this is probably because spontaneous and posed smiles differ in this aspect. In our data, the main difference was that spontaneous smiles tend to last longer than posed ones. Nevertheless, this might be influenced by the duration of the instruction to pose a smile, and the method used to elicit spontaneous expressions.

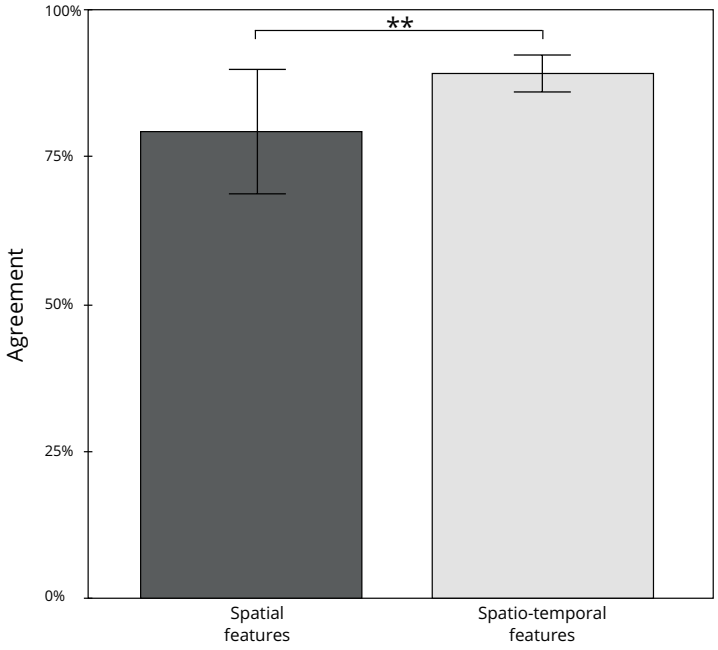
Moreover, the algorithm using spatial and magnitude EMG features achieved accuracy rates with more variability than the one using the spatio-temporal features. A possible explanation for this phenomenon is the ability of the participant to pose a smile. When the participants' posed and spontaneous smiles are visually very similar, the spatial-magnitude classifier has difficulties distinguishing them, hence, a low accuracy is achieved. This is the case of participant 9. On the other hand, when the participant is not able to produce a posed smile, the task becomes easier, and the accuracy increases. For example, the posed smile of participant 8 did not look like a smile. Then, the accuracy achieved reached 97.80%. With the spatio-temporal classifier, the accuracy for participant 9 is 85.23%, and for participant 8 is 92.74%. These values are higher in average, and less variable. Therefore, we might infer that people can train and learn how to move specific muscles to pose a smile, but controlling the timing of the changes is more difficult to achieve. Despite this, the spatio-temporal scores of Japanese nationals vary less than those of other nationalities. Thus, in a more balanced sample cultural differences might be observable.

## 4.3 The EMG signature of different smiles

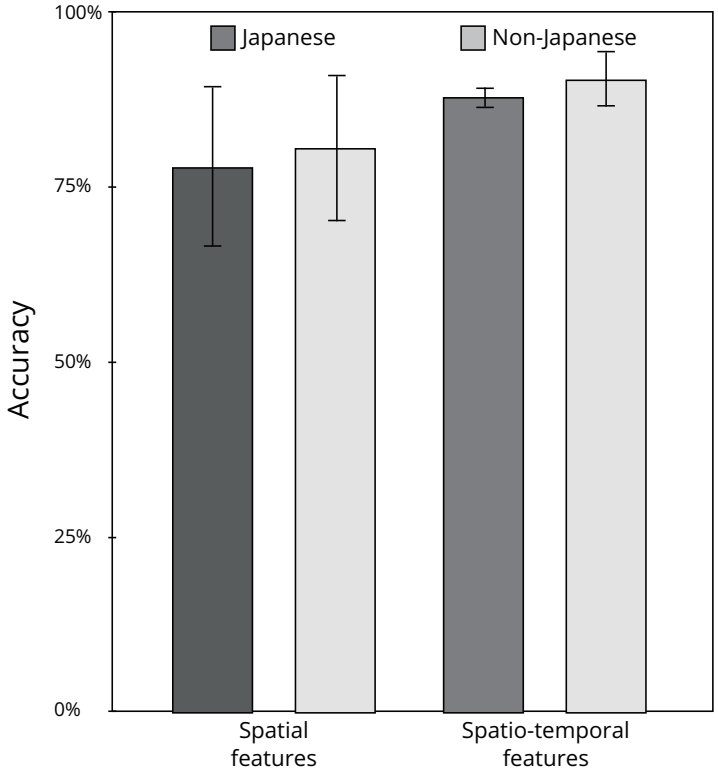
Although EMG has long been used to measure facial muscle activity, most of the work done so far implicated placing the electrode on top to the studied muscle [37]. This simple fact made it difficult for people to truly show spontaneous smiles. Thus, posed facial expressions were the most studied.

In this chapter, we used a wearable device that allows to measure distal facial EMG. This approach opens the possibility to unobtrusively track facial EMG signals of spontaneous smiles. Using the data collected from experiment 1, it was shown that distal EMG is an effective measure to identify fast and subtle spontaneous smiles, even at a micro-expression level (**RQ 1**). This method is specially advantageous when two or more people are being tracked (**RQ 1-b**), as occlusion and movement is common in this situation. Nevertheless, identifying and counting smiles is not enough to assess the private affective experience of a person. For that purpose, it is necessary to distinguish between posed and spontaneous smiles. Further data analysis showed that it is possible to identify the differences between both types of smiles using EMG (**RQ 2**). Whilst the spatial distribution of the muscles used to make these smiles differ, it was found that temporal features are more robust to distinguish between posed and spontaneous smiles. Specifically, smile duration, rising time, and decaying time (**RQ 2-b**).

Despite the good results achieved, there are some limitations in this study. First of all, micro-smiles are rarely displayed. This implies that their dynamics might differ from other smiles. Moreover, the unavailability of micro-smile examples causes a significant unbalancing in the data. This, in turn, might bias the performance of the machine learning algorithm. This limitation was carefully addressed by under-sampling the no-expression data. Another possible confounding is that the duration of the smiles might be constrained by the elicitation method in itself. Therefore, the next chapter describes a second experiment in which these limitations were addressed to obtain high-quality data.




(a) *Achieved accuracy.*




(b) *Achieved accuracy per nationality.*

Figure 4.8: *Performance differences among algorithms.* The Y-axis represents the accuracy in percentage. The average is presented for explanatory purposes. The X axis represents the category.

# Chapter 5

# Spatial and temporal dynamics of affective responses during spontaneous and posed smiles

In the previous chapter, the feasibility of detecting spontaneous smiles from EMG in an unobtrusive manner was proven. Furthermore, a first attempt to distinguish between posed and spontaneous smiles was made. However, in the previous experiment, posed and spontaneous smile duration could have been affected by the elicitation method used. Since posed expressions were elicited on command, their temporal dynamics might have been constrained. Therefore, experiment 2 was designed to collect posed smiles on a different paradigm.

Particularly, special care was taken to elicit facial expressions with known ground truth and without modifying their temporal characteristics. Therefore, a triple-check-method was used to ensure the validity of the ground truth labels. First, the experiment was carefully designed to create an appropriate valence in each block. A positive valence to elicit spontaneous smiles, and a slightly negative valence while participants were asked to pose smiles. Second, participants self-reported their emotions after each block, to confirm the induced affective state. Finally, participants were asked to video code their own facial expressions and report whether the smiles they made were posed

or spontaneous. In this manner, the validity of the ground truth label was ensured.

Moreover, other measurements were taken to further investigate affective responses during these two types smiles. Electrodermal Activity (EDA) was recorded as a measure of arousal, as well as head movement measured with an Inertial Measurement Unit (IMU). As mentioned in section 2.1, EDA complements the dimensional representation of emotion. On the other hand, as argued in section 2.3, head movement is hypothesized to increase accuracy when distinguishing between posed and spontaneous smiles.

Finally, the participants were sampled equally from three different cultural backgrounds to explore any differences among them. These were Japanese, Chinese and Brazilian. Furthermore, care was taken to maintain gender balance on the sample.

## 5.1 Experiment 2

### 5.1.1 Participants

38 voluntary participants took part on the study (19 female, average age=25.03 years old, SD=3.83). Participants were sampled from three nationality groups. 12 were Japanese (6 female), 13 were Chinese (7 female), and 14 were Brazilian (6 female). Since the experiment was conducted in Japan, the Chinese and Brazilian people were not in their native environment.

### 5.1.2 Experiment design

The experiment design consisted of six blocks (figure 5.1). The first block, from now on called "Spontaneous Block" (S-B), was aimed to induce a positive affective state, and therefore, elicit spontaneous smiles. The second block, "Neutral Block" (N) was aimed to revert the positive affect elicited during S-B into neutral affect. In the third block, participants were asked to pose a smile for the camera during approximately 5 s (P). In the fourth block, or the "Posed Block" (P-B), participants

were requested to "make similar facial expressions as you did when you watched the first video. Exactly, the ones 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 he cannot guess what video were you watching, then you are a good actor. Please do your best to win against the evaluator". This type of smile was considered as a posed smile that deliberately intends to convey the impression of having fun to a third person. The fifth block provided an opportunity to smile for the camera, again during 5 s (F-P). Finally, the last block was labeled as the moment when the experimenter informed the participant that the experiment was over. A smile was expected at that point as a sign of happiness because they finished all tasks (F-S).

All participants went through all experiment blocks in the same order. This was to keep the purpose of the experiment hidden during the spontaneous block. Only the stimuli videos inside the spontaneous block were counterbalanced. After both the S-B and the P-B, participants were asked to respond to affect-assessments via questionnaires. Furthermore, they were asked to video code their own facial expressions.

This research was approved by the Engineering Ethical Committee of the University of Tsukuba with review code 2017R176.

### 5.1.3 Stimuli

During the Spontaneous, Neutral, and Posed Blocks, 90 s videos were used as stimuli. Each stimuli video was preceded by a 10 s neutral video aimed to establish a relaxing baseline. The contents of the videos were the following:

- **Pre-block stimuli.** This video consisted of raindrops falling on the camera lenses.

- **Spontaneous Block.** Three funny and cute videos of 30 s were concatenated with a 1 s black transition. These were popular internet videos featured a baby getting surprised with a simple magic trick (from the previous experiment); a

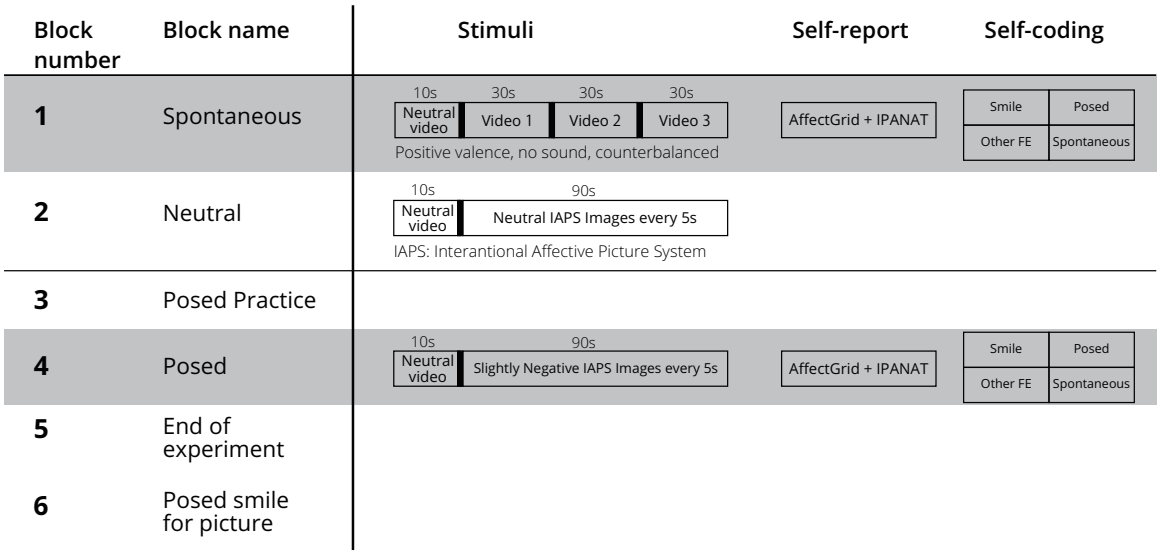



Figure 5.1: *Experimental design for experiment 2.* All participants went through all six experiment blocks in the same order. The first block was aimed to induce positive affect and therefore smiling behavior. The second block was aimed to reset that affective valence. The third block provided an opportunity to practice a posed smile when smiling for the camera. The fourth block aimed to induce a slightly negative feeling while people were deliberately asked to smile. The fifth block provided an opportunity to smile for the camera. Finally, the last block was labeled as the moment when the experimenter informed the participant that the experiment was over. A smile was expected at that point.

panda calling for the attention of the zoo guard [118]; and a cat moving in an interesting manner as his owner petted it [119]. These were intended to match the preferences of most of the participants. The three videos were presented in a counterbalanced order included all six orders.

- **Neutral Block.** The neutral block video consisted of 18 pictures from the International Affective Picture System (IAPS) [120] presented every 5 s, for a total of 90s. Hence, the duration of the neutral video was the same as the duration of the spontaneous block. These images were chosen to have a rated likeability from five to six points.

- **Posed Block.** Similarly to the neutral block, 18 IAPS pictures were selected and presented every 5s for a total of 90s. The images chosen for this block


Figure 5.2: *Wearable EMG channel position.* The EMG wearable consists of four channels located as depicted in this figure. EMG activity on channels 1 and 2 was significantly different between the spontaneous and the posed blocks.

had a midly unpleasant valence. The likeability of the images was constrained between four and five points.

### 5.1.4 Measurements

Four channels of distal facial EMG were measured from sides of the face using a Biolog DL4000 in the same wearable as in previous experiments. Figure 5.2 shows the EMG channel position on the wearable. EDA was measured from both the left hand index and ring fingers, and from the neck. Heart rate was measured using Photoplethysmography (PPG) sensors placed both on the middle finger of the left hand and the left earlobe. Head movement was measured with an Inertial Measurement Unit (IMU) placed on the back of the head. Additionally, an IMU recorded hand movements to aid motion artifact removal. EDA, PPG, and IMU measures were gathered using two Shimmer3 GSR+ units. All sensors were synchronized to the start of the stimuli.

Video of the participant's facial expressions was recorded using two cameras. One a Canon Ivis 52 at 30FPS, and the second a Intel RealSense camera at @60FPS, depth at 480x360, and color at 640x480 resolution. Furthermore, voice was recorded using two channels of a RASP-LC MEMS microphone array.

Participants were asked to answer the Affect Grid[13] as a measure of explicit affect

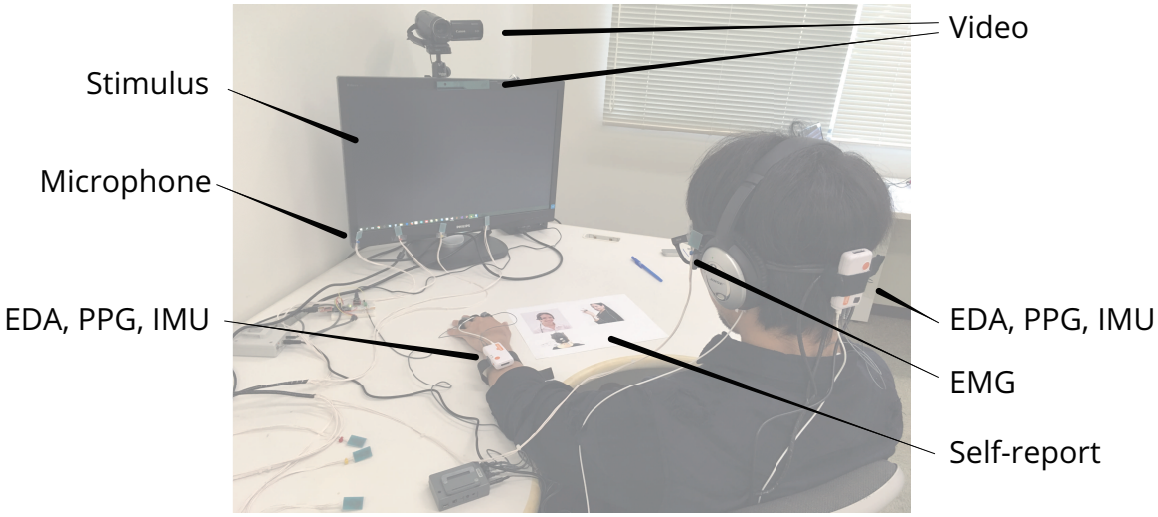



Figure 5.3: *Experiment setup for experiment 2.* All sensors are labeled in the picture to show their respective location.

self-report in a dimensional space; and the Implicit Positive and Negative Affect Test (IPANAT) [79, 121] as a measure of implicit affect. Additionally, they were asked to rank the videos in their order of preference after the spontaneous block, and to report if they had seen the videos before or if they would watch them again. At the end of the experiment, control questions about age, gender, hometown, and how often do they smile in everyday life were asked.

Finally, participants were also asked to tag the onset and offset of their own facial expressions, as well as labeling them as spontaneous or posed using the Dartfish Version 3.2 software with a customized set of label buttons.

### 5.1.5 Apparatus

All stimuli were presented to the participant in a Philips B-line 240B4 24 inches monitor with a resolution of 1920 x 1200 pixels. A MSi GP602PE230 Laptop was used to present the stimuli and to control the triggers to synchronize all devices. This Laptop was connected via Bluetooth to the Shimmer sensors used to record EDA, PPG, and IMU measurements. A wired connection was used to communicate with the RealSense camera and with the Display. Additionally, the stimuli laptop

Figure 5.4: *Wearables experiment 2.* From left to right, the EMG wearable, the usage of the Shimmer3 GSR on the hand, and the Shimmer3 GSR integrated on the EMG wearable to measure skin conductance from the neck.

was connected via USB to a custom hardware circuit. This circuit received wireless signals from a remote controller used by the experimenter to start the stimuli. Once the stimuli was started, a hardware trigger was sent to the Biolog hardware recording the EMG. Furthermore, a software trigger was inserted in the Shimmer data. This software trigger was sent using a local host UDP connection to the recording software. Finally, another two laptops were used. The first one, a Dell Latitude E6230, to record the EMG data received from the Biolog device via Bluetooth. The second, a Dell Inspiron N5110, to let the participants do their own video-coding using the Dartfish software.

### 5.1.6 Procedure

The experiment and questionnaires were conducted in Japanese, Chinese, and English. The participants chose their more comfortable language. Participants were invited to a "video rating" experiment. They were told that the goal was to rate the content of the videos and how comfortable was to wear the sensing wearables. They were informed that the experiment consisted of several blocks, and that the instructions for each block would be provided before starting each one. If they agreed to participate, they were asked to sign an informed consent. Next, they were shown

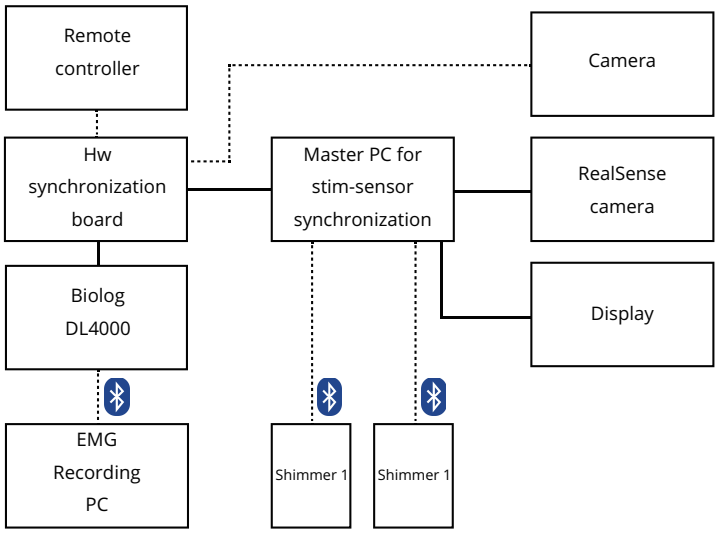



Figure 5.5: *Sensor synchronization experiment 2.* All sensors used during the experiment were synchronized to the start of the stimuli. A button on a remote controller triggered the stimuli presentation, a hardware trigger that marked the EMG, and displayed an LED light next to the participants face. Simultaneously, a software trigger was inserted into the Shimmer-recorded data using a local host UDP connection.

a picture with the wearables and aided by the experimenter to wear them. As mentioned in the experimental design section (5.1.2), the experiment always started with the spontaneous block. Participants were asked to watch the videos and relax. After the stimuli, they were explained how to answer the Affect Grid, the IPANAT questionnaire, and the video preference questions. While the participant was answering the questions, the experimenter copied and prepared for video-coding the videos with the participant's face. Next, participants were asked to tag any facial expressions they had made during the previous block. The tags included starting and end time of the expression, whether the expression was a smile or not, and if the expression was a posed expression or a spontaneous one. All participants were allowed to practice with a one minute video. After the video-coding, the neutral block started. Participants were told to watch the video and relax. After the neutral block, they were asked to smile for the camera as long as the screen changed to orange (5 s), in order to check the sensors' recordings. This was intended as practice for the posed smiles. Afterwards, participants were debriefed. An explanation about the goal of distinguishing

between posed and spontaneous smiles was provided. During the posed block, they were asked to watch the stimuli and to perform the same smiles they did during the spontaneous block as much as possible. Their objective was to make it impossible for a third person to tell the difference between the experimental blocks from the facial expressions alone. After watching the video, they answered the Affect Grid, and the IPANAT. The demographics and control questions were included afterwards. Next, they were asked to video code their own expressions as before. When they finished, all systems were still recording. Then, the experimenter informed them that it was the end of the experiment while inserting a trigger to mark that moment in the data. The experimenter then waited to see the reaction of the participant. After, the experimenter requested one last smile for the camera, while the screen changed to orange. Then the experiment was over, all recording systems were switched off, and the experimenter proceeded to remove all the wearables.

### 5.1.7 Analysis and Results

#### Self-report

A one-factor ANOVA revealed no significant self-perceived differences in how much participants smile ($F(2,36)=1.09$, $p > 0.05$, figure 5.6).

A 3-factor mixed ANOVA with the Affect Grid valence as the dependent variable, yielded significant results only for experiment block ($F(1,64)=11.465$, $p < 0.01$). Nationality ($F(2,64)=2.229$, $p > 0.05$), and gender ($F(1,64)=0.442$, $p > 0.05$) differences were not significant. The only significant interaction effect was the nationality and block interaction ($F(2,62)=3.754$, $p < 0.05$). Figure 5.7(a) depicts the differences in valence ratings between the posed and spontaneous experimental blocks, and between different nationalities. Participants reported more positive feelings during the spontaneous block, than during the posed block.

A similar ANOVA using the Affect Grid arousal as the dependent variable showed no significant differences in nationality ($F(2,64)=1.268$, $p > 0.05$), experiment block ($F(1,64)=0.498$, $p > 0.05$), nor gender ($F(1,64)=0.170$, $p > 0.05$). The interaction ef-

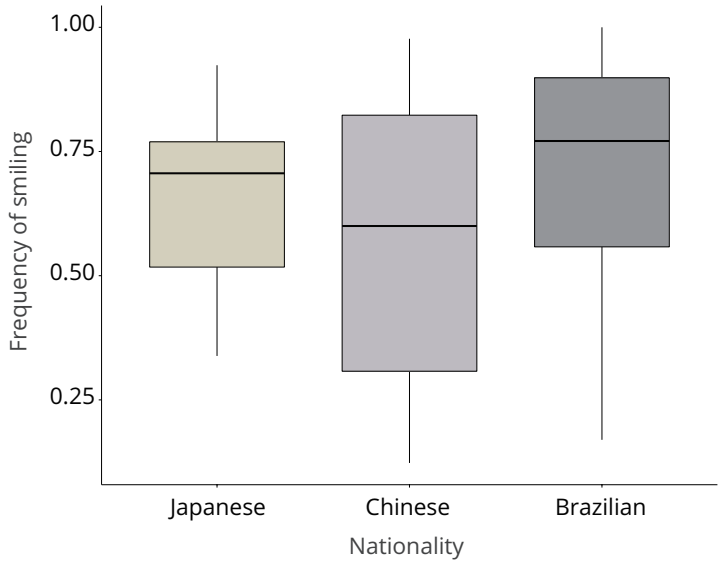


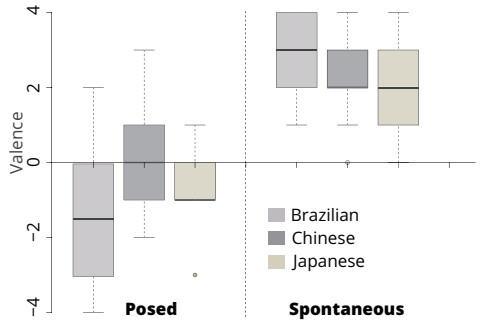

Figure 5.6: *Self-perceived smiling frequency.* When asked "How often do you smile?", participants of different nationalities did not report any significant difference in the frequency of smiling.

fects were also not significant. Figure 5.7(b) shows the differences in arousal ratings between the posed and spontaneous experimental blocks, and between different nationalities. Although the differences among experimental blocks are not significant, the high variability in the ratings is notorious. Specially for Brazilian and Japanese nationals.

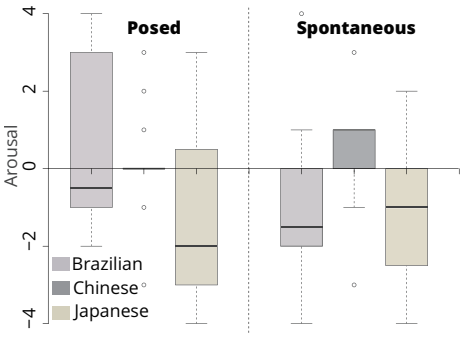
On the other hand, the IPANAT scores on positive and negative effect are shown on figure 5.8. A 3-factor mixed ANOVA with the IPANAT scores as dependant variable, yielded significant differences in the nationality ($F(2,140)=4.389$, $p < 0.05$). However, experiment block ($F(1,140)=0.023$, $p > 0.05$), and reported affect ($F(1,140)=0.104$, $p > 0.05$) differences were not significant. Interaction effects were not significant.

### Video coding

According to their own video coding, 272 smiles were elicited from 32 participants. 127 were spontaneous, and 145 were posed. Only three people produced sounds that would be cataloged as laughter. According to the participants comments during the video-coding part of the experiment, it is difficult to know if a smile was spontaneous




(a) *The effect of the experiment block and nationality on valence.* Participants reported a more positive valence during the spontaneous block, and less positive valence during the posed block.




(b) *The effect of the experiment block and nationality on arousal.* Participants ratings of arousal during the posed block were not different form each other.

Figure 5.7: *Affect grid ratings per experimental block and nationality* The average ratings of valence and arousal per experimental block and nationality are shown.

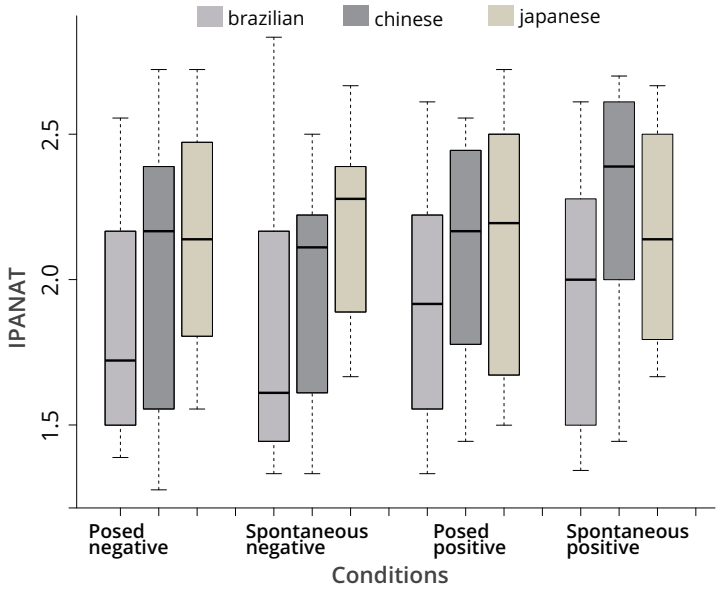



Figure 5.8: *IPANAT scores per nationality, experiment block, and reported affect.* The IPANAT scores are similar for both spontaneous and posed blocks.

or posed even for themselves. Specially Brazilians mentioned that sometimes a posed smile transformed into a spontaneous one when they thought about the irony of having to smile to the conflicting stimuli images.

Table 5.1: *Inter-coder agreement.* The agreement between the different coders is shown. A considerably good agreement is achieved when judging between smiles and other expressions. However, the agreement decreases when the judgment also considers whether the facial expression is posed or spontaneous.

| | Smile-No smile agreement | Posed-Spontaneous agreement |
|---|---|---|
| Cohen's Kappa between independent coders | 0.5880 | 0.2951 |
| Fleiss' Kappa between the two coders and the participant | 0.5673 | 0.1299 |

Besides the participant's own video coding, two independent raters labeled the videos. The two independent coders used the same software as the participants (Dartfish Version 3.2). They coded for the start frame and the duration of every facial expression. They labeled each expression as a smile, or another facial expression; and as a posed or spontaneous expression. Additionally, they labeled the involved FACS Action Units (AU). The coded AU were AU1, AU02, AU04, AU05, AU6, AU09, AU10, AU12, AU14, AU15, AU17, AU18, AU25, AU26, AU28, and AU38. Smiles were often a display of AU6 and/or AU12. However, the smile label was not assigned every time these AU occurred. They Coder 1 was familiar with the experimental design, whilst Coder 2 was not.

When discriminating whether participants were smiling or not, the Cohen's Kappa for agreement between the two independent coders was 0.588. In the same task, the Fleiss' Kappa between the two coders and the participant's own video coding was 0.567. However, the agreement diminished when the task was to determine whether the expressions displayed were posed or spontaneous. The posed-spontaneous Cohen's Kappa agreement between the two independent coders was 0.295. When also including the own coding from the participant, the Fleiss' Kappa scored 0.129. These results are summarized in table 5.1.

### Electromyography

A similar algorithm to the one described on section 4.2.2 was used to calculate the temporal features of different smiles. First, an envelope is fitted to the rectified EMG Independent Components (IC). Then the maximum and minimum points of the envelope are determined. Based on those peaks, maximum magnitude, rising time, decaying time, rising speed, decaying speed, and duration of the smile are calculated. A series of t-tests between posed and spontaneous features revealed that only rising time ($t(611) = -2.0859$, $p < 0.05$) and decaying speed ($t(575) = -2.5122$, $p < 0.05$) differences were significant.

Additional analysis were performed on the filtered EMG data. The applied filters were a notch filter at 50 Hz, and a band-pass filter from 5 to 350 Hz. Then, the EMG value was averaged over each block, and divided by the duration of that block. These ratio values were calculated per participant and per EMG channel. A mixed factor ANOVA showed a significant difference between this EMG ratio values in nationality ($F(2,280)=6.536$, $p < 0.01$), experimental block ($F(1,280)=7.072$, $p < 0.01$), and channel ($F(3,280)=3.635$, $p < 0.01$). Additionally, the interaction effect between Nationality and Block was significant ($F(2,280)=6.461$, $p > 0.05$). Given this difference, a repeated measures ANOVA with experimental block, and Independent Component (IC) as independent variables, and the magnitude of each smile was performed. No significant differences were found. Neither in experimental block ($F(1,439)=0.185$, $p > 0.05$), or IC ($F(1,439)=0.010$, $p > 0.01$).

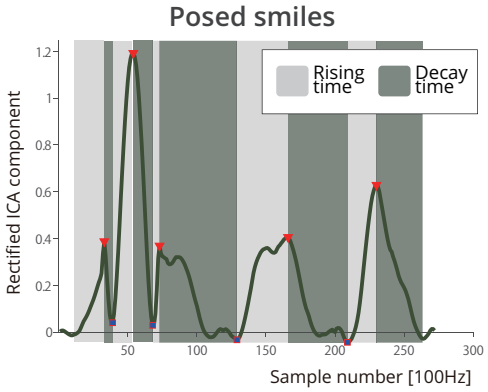
### Electrodermal activity

EDA responses mildly fluctuated with the stimuli blocks. Hand EDA changed more than neck EDA. It displayed typical tonic and phasic changes. On the neck EDA, phasic changes were about for times smaller in magnitude (figure 5.10).

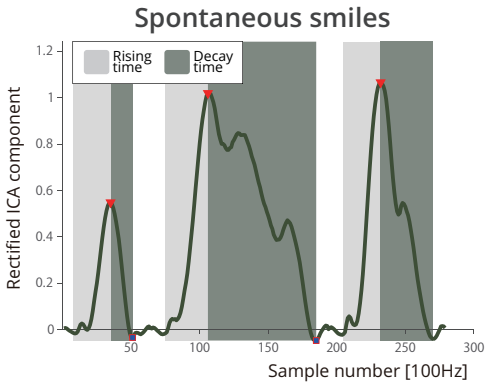
Two types of analysis were performed to investigate the differences between EDA events co-occurring with posed and spontaneous smiles. The first one was to investigate the potential of using the data gathered to automatically distinguish between

Table 5.2: *EMG spatio-temporal features-based posed and spontaneous smiles identification for experiment 2.* Spatio-temporal features were used as described in section 5.1.7. The high accuracy from the previous experiment was maintained at around 91% with a standard deviation of 4%. The results of only 27 participants are shown, as not every participant smiled during the experiment.

| Participant | Precision | Recall | Accuracy | Number Posed Smiles | Number Spontaneous Smiles | Nationality | Gender |
|---|---|---|---|---|---|---|---|
| 1 | 90.00% | 96.18% | 90.78% | 5 | 3 | Japanese | Male |
| 2 | 81.58% | 91.18% | 87.65% | 5 | 5 | Japanese | Male |
| 3 | 83.50% | 100.00% | 86.92% | 5 | 4 | Japanese | Female |
| 5 | 90.00% | 98.02% | 90.91% | 7 | 4 | Japanese | Male |
| 7 | 100.00% | 100.00% | 100.00% | 4 | 3 | Brazilian | Female |
| 8 | 100.00% | 50.00% | 95.45% | 3 | 2 | Chinese | Female |
| 9 | 90.00% | 100.00% | 93.94% | 4 | 3 | Japanese | Male |
| 10 | 90.24% | 92.50% | 90.14% | 5 | 4 | Brazilian | Female |
| 11 | 88.89% | 95.24% | 91.46% | 5 | 4 | Japanese | Male |
| 13 | 86.49% | 94.12% | 88.71% | 3 | 2 | Brazilian | Male |
| 14 | 89.71% | 100.00% | 91.03% | 3 | 2 | Japanese | Female |
| 17 | 89.71% | 100.00% | 100.00% | 9 | 5 | Brazilian | Female |
| 18 | 94.34% | 86.21% | 90.00% | 8 | 4 | Chinese | Male |
| 19 | 100.00% | 71.43% | 90.00% | 2 | 3 | Brazilian | Male |
| 20 | 100.00% | 85.00% | 92.86% | 2 | 3 | Japanese | Female |
| 21 | 83.33% | 93.75% | 88.89% | 3 | 2 | Chinese | Male |
| 23 | 92.45% | 98.00% | 93.98% | 5 | 4 | Japanese | Female |
| 24 | 95.95% | 98.61% | 95.51% | 11 | 3 | Chinese | Female |
| 26 | 89.47% | 58.62% | 82.50% | 10 | 7 | Chinese | Female |
| 29 | 100.00% | 28.57% | 87.50% | 1 | 3 | Chinese | Female |
| 31 | 90.00% | 100.00% | 91.67% | 4 | 2 | Chinese | Female |
| 34 | 89.66% | 100.00% | 90.63% | 5 | 2 | Chinese | Male |
| 35 | 87.76% | 89.58% | 90.00% | 5 | 8 | Brazilian | Female |
| 36 | 91.57% | 100.00% | 93.58% | 12 | 7 | Brazilian | Male |
| 37 | 83.78% | 96.88% | 86.54% | 3 | 4 | Brazilian | Male |
| 39 | 95.00% | 95.00% | 93.02% | 8 | 3 | Brazilian | Male |
| 41 | 88.24% | 62.50% | 92.57% | 5 | 7 | Brazilian | Female |
| Average | 91% | 88% | 91% | 5.26 | 3.81 | | |
| SD | 5% | 18% | 4% | 2.74 | 1.68 | | |




(a) *Posed smile envelope example.* The processed EMG envelope of self-reported posed smiles during the posed block are shown. This is an example for participant 37, Independent Component number 1.




(b) *Spontaneous smile envelope example.* The processed EMG envelope of self-reported spontaneous smiles during the spontaneous block are shown. This is an example for participant 37, Independent Component number 1.

Figure 5.9: *EMG envelopes from posed and spontaneous smiles.* Although both types of smiles are visually different, only the rising time and decaying speed are significantly different.

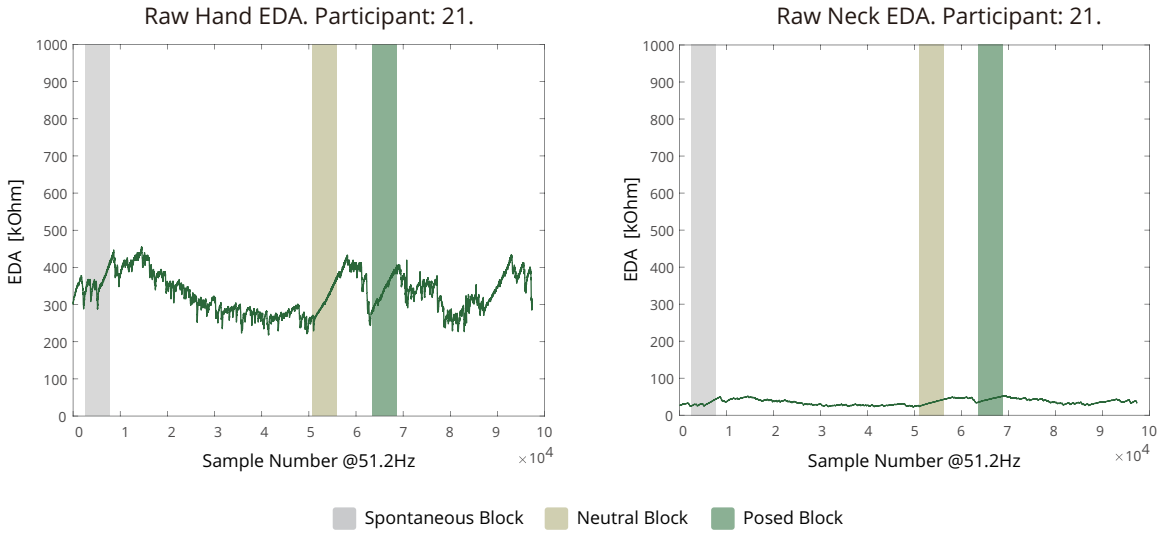



Figure 5.10: *Skin conductance from the hand and neck.* The skin conductance measured from participant 21 is shown in the figure. There are similar trends in both measurements, but the degree of movement for the neck EDA is much less than for the hand.

spontaneous and posed events. This was done both on the neck and hand data. Second, the differences in the EDA responses per experimental condition, nationality,

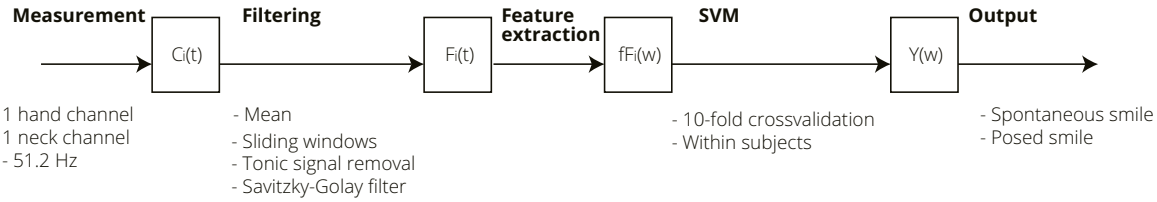
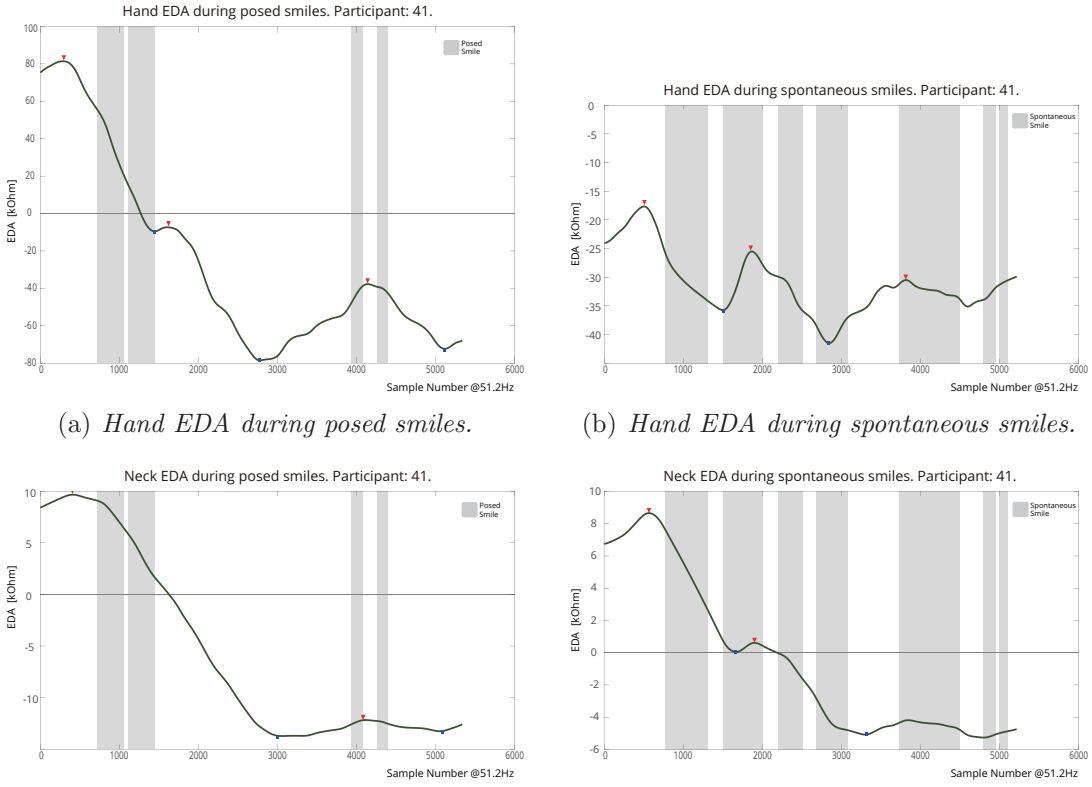



Figure 5.11: *Skin conductance signal processing.* The skin conductance measured from both the neck and the hand of each participant was first smoothed by using a 100 ms sliding window overlapping each sample. Then, the tocnic singal was removed by subtracting the two coarsest coefficients from a Discrete Cosine Transform from the original signal. Finally, a Savitzky-Golay Filter was used to remove motion artifact peaks. Then this signal was labelled and used to extract relevant features to train a SVM to distinguish between posed and spontaneous events.

gender, and electrode location were explored to identify differences. For both types of analysis the same pre-processing procedure was used.

**Pre-processing.** The skin conductance measured from both the neck and the hand of each participant was first smoothed by using a 100 ms sliding window overlapping each sample. Then, the tocnic singal was removed by subtracting the two coarsest coefficients from a Discrete Cosine Transform from the original signal [44]. Finally, a Savitzky-Golay Filter with a 1st order polynomial and 1001 as frame length was used to remove motion artifact peaks (figure 5.11). The Savitzky-Golay Filter parameters were selected by visual inspection. The selection criteria was to remove motion-related artifacts in the EDA signal.

Figure 5.12 The plots show EDA responses measured from the hand and the neck during posed and spontaneous smiles for participant 41. The green line shows the pre-processed EDA signal, and vertical bars show the smile labels. In general, Hand EDA fluctuates more than Neck EDA. Albeit different, they seem to correlate. Visual inspection of the EDA plots suggests that EDA peaks anticipate smiles. However, this is not always the case. For several participants, the EDA during posed smiles present very little phasic changes. On the other hand, spontaneous reactions are characterized by frequent phasic changes.




(a) *Hand EDA during posed smiles.* (b) *Hand EDA during spontaneous smiles.*

(c) *Neck EDA during posed smiles.* (d) *Neck EDA during spontaneous smiles.*

Figure 5.12: *Labeled EDA from the hand and neck.* The plots show EDA responses measured from the hand and the neck during posed and spontaneous smiles for participant 41. The green line shows the pre-processed EDA signal, and vertical bars show the smile labels. In general, Hand EDA fluctuates more than when Neck EDA. Albeit different, they seem to correlate.

**Feature extraction.** As suggested by [122, 123], a set of features were extracted from the EDA. The magnitude ratio of the absolute value of the EDA signal and the smile duration; the mean of the first order derivative of the EDA signal per smile; and the number of peaks divided by the minimum smile width. This data is considered as the first training set. A drawback of using these features, is that the amount of information available is reduced by the number of smiles elicited. Since not every participant smiled, and many smiled very little, other features were considered.

The magnitude of the pre-processed EDA signal during smile episodes was used to increase the amount of data points available for training. These two feature sets used to train a SVM. Additionally, the average magnitude of the EDA signal and its

Table 5.3: *EDA-based identification using peak features.* The magnitude ratio of the absolute value of the EDA signal and the smile duration; the mean of the first order derivative of the EDA signal per smile; and the number of peaks divided by the minimum smile width were used as features to train a SVM. A limited number of participants smiled frequently enough to compute those features.

| Participant | Precision | Recall | Accuracy | Number Posed Smiles | Number Spontaneous Smiles |
|---|---|---|---|---|---|
| 3 | 80% | 100% | 88% | 5 | 4 |
| 5 | 100% | 100% | 100% | 7 | 4 |
| 17 | 83% | 100% | 90% | 9 | 5 |
| 18 | 100% | 100% | 100% | 8 | 4 |
| 23 | 100% | 100% | 100% | 5 | 4 |
| 26 | 100% | 71% | 86% | 10 | 7 |
| 36 | 100% | 83% | 92% | 12 | 7 |
| 41 | 88% | 100% | 93% | 5 | 7 |
| Average | 93.9% | 94.4% | 93.5% | 7.6 | 5.3 |
| SD | 8% | 10% | 5% | 2.45 | 1.39 |

first order derivative divided by the experimental block duration were also calculated. These last features were used as a measure to compare whether there is a significant difference between EDA responses per block, nationality, and gender.

**Identification of EDA responses during posed and spontaneous smiles.** Three models were trained using a Support Vector Machine (SVN) with a Gaussian Kernel Function in a cross-validation with 70% train, 15% validation, and 15% test data partition. Participants who did not have enough training data (12, 15, 16, 30 and 32) were excluded. Another two participants (2 and 6) were excluded due to error during the measurement of the EDA data. The first model used the first set of features. The second and the third models used the magnitude of the EDA signal from the neck and the hand, respectively. The results are detailed in tables 5.3, 5.4 and 5.5.

**Comparison between experimental blocks, nationalities, and gender.** Four-factor mixed design ANOVAs were performed. Independent variables were nationality, experimental block, location of the EDA sensor, and gender. The dependent variables were: (1) the average magnitude divided by the total duration of the block; (2) the average magnitude of the EDA signal per block; and (3) the average mag-

Table 5.4: *Experiment 2 identification results using hand EDA.*

| Participant | Precision | Recall | Accuracy | Number Posed Smiles | Number Spontaneous Smiles | Nationality | Gender |
|---|---|---|---|---|---|---|---|
| 1 | 100% | 100% | 100% | 5 | 3 | japanese | Male |
| 3 | 93% | 94% | 90% | 5 | 4 | japanese | Female |
| 4 | 100% | 100% | 100% | 1 | 1 | japanese | Female |
| 5 | 100% | 100% | 100% | 7 | 4 | japanese | Male |
| 7 | 100% | 100% | 100% | 4 | 3 | brazilian | Female |
| 8 | 100% | 100% | 100% | 3 | 2 | chinese | Female |
| 9 | 78% | 86% | 92% | 4 | 3 | japanese | Male |
| 10 | 100% | 100% | 100% | 5 | 4 | brazilian | Female |
| 11 | 100% | 90% | 97% | 5 | 4 | japanese | Male |
| 13 | 100% | 100% | 100% | 3 | 2 | brazilian | Male |
| 14 | 100% | 100% | 100% | 3 | 2 | Japanese | Female |
| 17 | 100% | 100% | 100% | 9 | 5 | Brazilian | Female |
| 18 | 100% | 100% | 100% | 8 | 4 | Chinese | Male |
| 19 | 100% | 100% | 100% | 2 | 3 | brazilian | Male |
| 20 | 85% | 100% | 87% | 2 | 3 | Japanese | Female |
| 21 | 100% | 100% | 100% | 3 | 2 | Chinese | Male |
| 22 | 100% | 100% | 100% | 0 | 1 | Brazilian | Male |
| 23 | 88% | 87% | 86% | 5 | 4 | Japanese | Female |
| 24 | 100% | 100% | 100% | 11 | 3 | Chinese | Female |
| 25 | 100% | 100% | 100% | 0 | 1 | Mexican | Male |
| 26 | 89% | 91% | 87% | 10 | 7 | Chinese | Female |
| 27 | 100% | 100% | 100% | 1 | 2 | Japanese | Female |
| 28 | 100% | 100% | 100% | 0 | 4 | Chinese | Female |
| 29 | 100% | 59% | 84% | 1 | 3 | Chinese | Female |
| 31 | 100% | 100% | 100% | 4 | 2 | Chinese | Female |
| 33 | 100% | 100% | 100% | 0 | 4 | Chinese | Male |
| 34 | 100% | 100% | 100% | 5 | 2 | Chinese | Male |
| 35 | 76% | 38% | 86% | 5 | 8 | Brazilian | Female |
| 36 | 100% | 100% | 100% | 12 | 7 | Brazilian | Male |
| 37 | 100% | 100% | 100% | 3 | 4 | Brazilian | Male |
| 38 | 100% | 100% | 100% | 0 | 4 | Brazilian | Male |
| 39 | 89% | 99% | 90% | 8 | 3 | Brazilian | Male |
| 40 | 100% | 100% | 100% | 1 | 3 | Brazilian | Female |
| 41 | 95% | 88% | 96% | 5 | 7 | Brazilian | Female |
| Average | 97% | 95% | 97% | 4.12 | 3.47 | | |
| SD | 6% | 13% | 5% | 3.22 | 1.70 | | |

Table 5.5: *Experiment 2 identification results using neck EDA.*

| Participant | Precision | Recall | Accuracy | Number Posed Smiles | Number Spontaneous Smiles | Nationality | Gender |
|---|---|---|---|---|---|---|---|
| 1 | 100% | 100% | 100% | 5 | 3 | japanese | Male |
| 3 | 97% | 96% | 94% | 5 | 4 | japanese | Female |
| 4 | 100% | 100% | 100% | 1 | 1 | japanese | Female |
| 5 | 100% | 100% | 100% | 7 | 4 | japanese | Male |
| 7 | 100% | 100% | 100% | 4 | 3 | brazilian | Female |
| 8 | 100% | 100% | 100% | 3 | 2 | chinese | Female |
| 9 | 98% | 100% | 100% | 4 | 3 | japanese | Male |
| 10 | 100% | 100% | 100% | 5 | 4 | brazilian | Female |
| 11 | 100% | 100% | 100% | 5 | 4 | japanese | Male |
| 13 | 100% | 100% | 100% | 3 | 2 | brazilian | Male |
| 14 | 100% | 100% | 100% | 3 | 2 | Japanese | Female |
| 17 | 100% | 100% | 100% | 9 | 5 | Brazilian | Female |
| 18 | 100% | 100% | 100% | 8 | 4 | Chinese | Male |
| 19 | 100% | 100% | 100% | 2 | 3 | brazilian | Male |
| 20 | 80% | 99% | 80% | 2 | 3 | Japanese | Female |
| 21 | 100% | 100% | 100% | 3 | 2 | Chinese | Male |
| 22 | 100% | 100% | 100% | 0 | 1 | Brazilian | Male |
| 23 | 100% | 100% | 100% | 5 | 4 | Japanese | Female |
| 24 | 100% | 100% | 100% | 11 | 3 | Chinese | Female |
| 25 | 100% | 100% | 100% | 0 | 1 | Mexican | Male |
| 26 | 100% | 100% | 100% | 10 | 7 | Chinese | Female |
| 27 | 100% | 100% | 100% | 1 | 2 | Japanese | Female |
| 28 | 100% | 100% | 100% | 0 | 4 | Chinese | Female |
| 29 | 100% | 100% | 100% | 1 | 3 | Chinese | Female |
| 31 | 100% | 100% | 100% | 4 | 2 | Chinese | Female |
| 33 | 100% | 100% | 100% | 0 | 4 | Chinese | Male |
| 34 | 100% | 100% | 100% | 5 | 2 | Chinese | Male |
| 35 | 100% | 100% | 100% | 5 | 8 | Brazilian | Female |
| 36 | 100% | 100% | 100% | 12 | 7 | Brazilian | Male |
| 37 | 100% | 100% | 100% | 3 | 4 | Brazilian | Male |
| 38 | 100% | 100% | 100% | 0 | 4 | Brazilian | Male |
| 39 | 100% | 100% | 100% | 8 | 3 | Brazilian | Male |
| 40 | 100% | 100% | 100% | 1 | 3 | Brazilian | Female |
| 41 | 98% | 63% | 90% | 5 | 7 | Brazilian | Female |
| Average | 99% | 99% | 99% | 4.12 | 3.47 | | |
| SD | 3% | 6% | 4% | 3.22 | 1.70 | | |

nitude of the first order derivative of the EDA response in each block, divided by the duration of the block. None of the tests were significant, although in the three cases, the effect of the location from which the EDA was measured was marginally significant ($F(1,115)=3.519$, $p = 0.063$).

#### Head movement

Head movement was also measured. An embedded algorithm in the measuring device allowed to estimate the orientation of the Shimmers. The calculated Quaternion data was smoothed using a Savitzky-Golay Filter with a 1st order polynomial and 301 as frame length. Figure 5.13 shows an example of the resulting data per block condition and location. From these, no clear difference between the experimental blocks, and the smile type. However, when the magnitude data is used to train a SVM in a cross-validation with 70% train, 15% validation, and 15% test data partition. As shown in table5.6, the results are very good, and quite similar to the ones achieved by measuring EMG in the neck.

Finally, comparisons of the mean magnitude of the head orientation divided by the duration of the experiment block, per nationality, experiment block, and gender showed no significant differences. However, visual inspection of the plot suggests that participants moved more while smiling in the posed block. On the other hand, they moved more in between smiles in the posed block.

#### Comparison

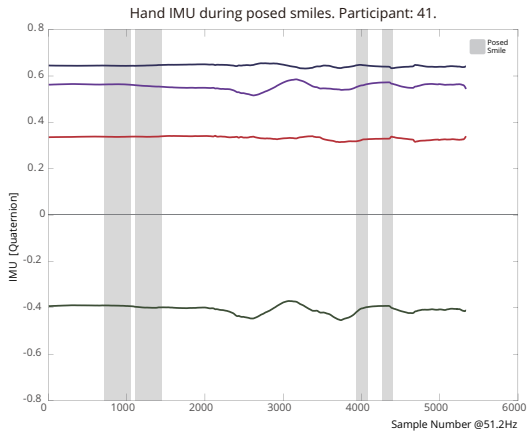
Figure 5.14 shows a plot comparing the accuracies achieved by each modality. The IMU-based classification achieved the best results, followed closely by the rest of the modalities.

### 5.1.8 Discussion

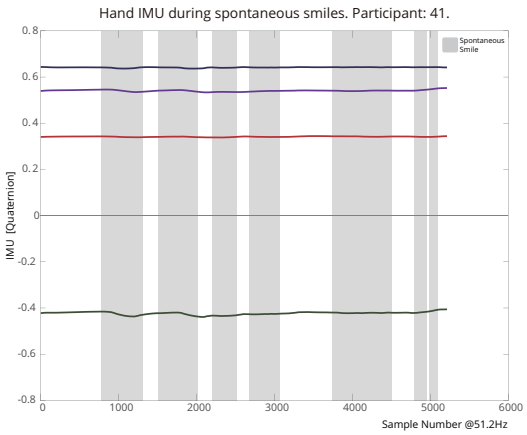
In this experiment, posed and spontaneous smiles were elicited and analyzed. Spontaneous smiles were elicited by showing positive-valenced videos to the partici-

Table 5.6: *Experiment 2 identification results using head IMU.*

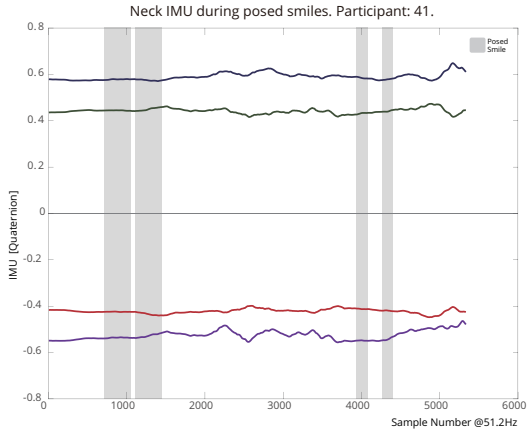
| Participant | Precision | Recall | Accuracy | Number Posed Smiles | Number Spontaneous Smiles | Nationality | Gender |
|---|---|---|---|---|---|---|---|
| 1 | 100% | 100% | 100% | 5 | 3 | Japanese | Male |
| 3 | 100% | 100% | 100% | 5 | 4 | Japanese | Female |
| 4 | 100% | 100% | 100% | 1 | 1 | Japanese | Female |
| 5 | 100% | 100% | 100% | 7 | 4 | Japanese | Male |
| 7 | 100% | 100% | 100% | 4 | 3 | Brazilian | Female |
| 8 | 100% | 100% | 100% | 3 | 2 | Chinese | Female |
| 9 | 100% | 100% | 100% | 4 | 3 | Japanese | Male |
| 10 | 100% | 100% | 100% | 5 | 4 | Brazilian | Female |
| 11 | 100% | 100% | 100% | 5 | 4 | Japanese | Male |
| 13 | 100% | 100% | 100% | 3 | 2 | Brazilian | Male |
| 14 | 100% | 100% | 100% | 3 | 2 | Japanese | Female |
| 17 | 100% | 100% | 100% | 9 | 5 | Brazilian | Female |
| 18 | 100% | 100% | 100% | 8 | 4 | Chinese | Male |
| 19 | 100% | 100% | 100% | 2 | 3 | Brazilian | Male |
| 20 | 100% | 100% | 100% | 2 | 3 | Japanese | Female |
| 21 | 100% | 100% | 100% | 3 | 2 | Chinese | Male |
| 22 | 100% | 100% | 100% | 0 | 1 | Brazilian | Male |
| 23 | 100% | 100% | 100% | 5 | 4 | Japanese | Female |
| 24 | 100% | 100% | 100% | 11 | 3 | Chinese | Female |
| 25 | 100% | 100% | 100% | 0 | 1 | Mexican | Male |
| 26 | 100% | 100% | 100% | 10 | 7 | Chinese | Female |
| 27 | 100% | 100% | 100% | 1 | 2 | Japanese | Female |
| 28 | 100% | 100% | 100% | 0 | 4 | Chinese | Female |
| 29 | 100% | 100% | 100% | 1 | 3 | Chinese | Female |
| 31 | 100% | 100% | 100% | 4 | 2 | Chinese | Female |
| 33 | 100% | 100% | 100% | 0 | 4 | Chinese | Male |
| 34 | 100% | 100% | 100% | 5 | 2 | Chinese | Male |
| 35 | 100% | 100% | 100% | 5 | 8 | Brazilian | Female |
| 36 | 100% | 100% | 100% | 12 | 7 | Brazilian | Male |
| 37 | 100% | 100% | 100% | 3 | 4 | Brazilian | Male |
| 38 | 100% | 100% | 100% | 0 | 4 | Brazilian | Male |
| 39 | 100% | 100% | 100% | 8 | 3 | Brazilian | Male |
| 40 | 100% | 100% | 100% | 1 | 3 | Brazilian | Female |
| 41 | 100% | 100% | 100% | 5 | 7 | Brazilian | Female |
| Average | 100% | 100% | 100% | 4.12 | 3.47 | | |
| SD | 0% | 0% | 0% | 3.22 | 1.70 | | |



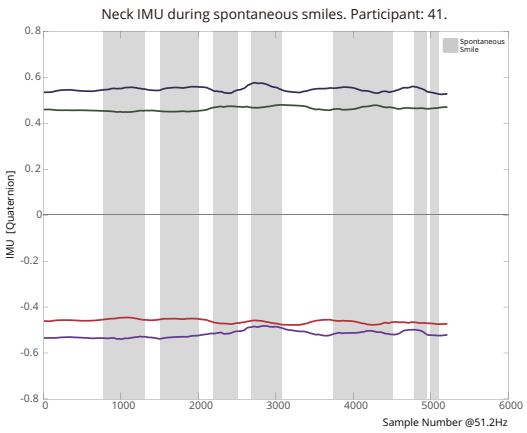

(a) *Posture from Hand IMU during posed smiles.*




(b) *Posture from Hand IMU during spontaneous smiles.*




(c) *Posture from Neck IMU during posed smiles.*




(d) *Posture from Neck IMU during spontaneous smiles.*

Figure 5.13: *Labeled IMU from the hand and neck.* The plots show head posture as measured from the hand and the neck IMU during posed and spontaneous smiles for participant 41. The color lines show the calculated quaternions, and vertical bars show the smile labels. It seems that participants moved more while smiling in the spontaneous block. On the other hand, they moved more in between smiles in the posed block.

pants. Posed smiles were asked, even during an slightly unpleasant situation. This experimental design allowed to some extent control the affect felt by the participants when they produced the required smiles. As a validation check, self-reported measures of affect were applied, and the participants themselves were asked to label their spontaneous and posed expressions.

From the self-reported measures, a valence difference between spontaneous and

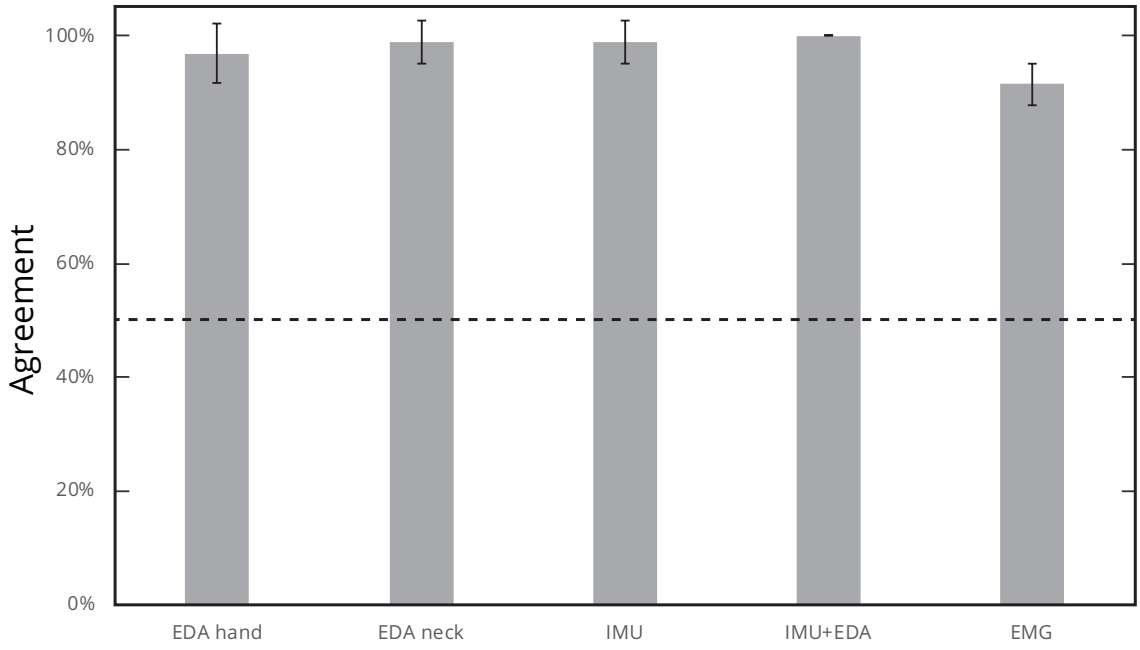



Figure 5.14: *Identification of posed and spontaneous responses per modality.*

posed blocks was observed. As expected, participants reported feeling more positive during the spontaneous block than during the posed block, independently from their nationality or gender. On the other hand, no arousal difference between spontaneous and posed blocks was observed. This might probably be because of the mildness of the video contents. They were pleasant enough for people to smile, but the intensity was similar among the videos of all blocks.

Similar to the arousal, the reported IPANAT scores presented no differences between experimental blocks. In this case, Japanese seemed to give higher scores for both positive and negative affect than the Brazilians. This might be due to the fact that in Japanese language, many contextual inferences are made. Therefore, it might be the case that Japanese nationals felt more confident assigning a meaning of the IPANAT words. However, these values were not consistent with the explicit self-report given in the Affect Grid. Explicit self-report (Affect Grid) and Implicit self-report (IPANAT) are not correlated, in neither block ($p > 0.5$). This seems to suggest that both measurement tools are measuring different levels of affective aware-

ness. This was previously observed by [16], who suggested that implicit self-report is related to autonomic affective responses, and independent from explicit self-report.

For the affective assessment to be done automatically, it is necessary to establish a ground truth. As explained in section 3.5, a machine can learn to agree to a predefined data label. These data labels can be assigned on a perceptual basis, considering only visible cues. However, a decision on whether the facial expression is posed or spontaneous, is an inference. Table 5.1 showed that the perceptual judgment, yields a higher agreement than the inferential one. Moreover, the agreement between independent coders and the participant's coding is even lower. These results suggests that, in this case, video-coding based only on visually perceivable cues is not the best method to establish the ground truth. Rather than relying on video-coding, a good experimental design when collecting the ground truth data is of utmost importance.

In the previous experiment results (section 4.2.3), the duration of the smiles did differ. As previously discussed and supported by the new results, it seems that the duration difference was influenced by the elicitation tasks. In this experiment we took special care to let the participants pose freely the smiles. The elicited smiles were considered as smiles that deliberately intend to convey the impression of having fun. Nevertheless, rising time was still significantly different between the two types of smiles. Furthermore, decaying speed turned out to significantly differ. As in the previous experiment, magnitude of posed and spontaneous smiles also did not differ. Moreover, high accuracy was maintained by using spatio-temporal features in this data set. This further supports the hypothesis that the difference among the smiles lies mainly on the temporal dynamics of the smiles.

Filtered EMG ratios showed that the EMG activity of the participants differs among experimental blocks, nationality, channel, and the interaction between nationality and block. However, there is not a clear reason why these differences are observed. It seems that the EMG activation from Chinese participants has a wider Standard Deviation than the other two nationality groups. Furthermore, the activity of channels one and two, placed on the right side of the face, is higher than those

placed on the left side. Previously, it was suggested that one of the features of enjoyment smiles are the symmetrical changes in the zygomatic major action on both sides of the face [37]. Therefore, this could be a reason for the asymmetry in the significant differences. Another explanation might have been measurement noise, as the effect disappears when the Independent Components are calculated. The magnitudes of the temporal features per IC do not significantly differ. Only the interaction between rising speed and IC. This could be because not all muscles involved move at the same speed.

Skin conductance on the hands changes more prominently than from the neck. However, both measures are correlated. The same algorithm was used to process both signal sources. An effort was made to reduce motion artifacts, and different features were calculated to train a SVM to classify between posed and spontaneous events. The results of this algorithm were very good. However, commonly used features such as magnitude, first order derivatives, and peaks per smile have the disadvantage that only one feature can be calculated per smile. This is because the EDA responses are too slow to happen within the smile samples. Hence, magnitude of the EDA alone was preferred to increase the data available for training and testing. The success of the magnitude features alone compared to other features might be due to the amount of data available.

Despite the magnitude differences between neck and hand EDA, both were proven to be suitable for the classification task. The differences in their magnitudes were only marginally significant. From visual inspection of the measured EDA responses, it seems that EDA changes anticipate facial movement. This was already suggested in one of the studies mentioned by [62]. However, further analyses are needed to confirm this hypothesis.

Moreover, the obtained EDA results have to be considered carefully. Although the pre-processing was chosen to remove motion artifacts. The effectiveness of the method was only evaluated by visual inspection. Furthermore, the accuracy obtained by using motion sensing is higher than that of EDA. This suggests that motion can

explain the differences between spontaneous and posed affective responses quite well. Therefore the covariance of EDA and IMU data might explain the predictive power of EDA, even if participants did not explicitly report to be aroused. Furthermore, the differences in magnitudes between blocks were non significant.

Surprisingly, the head IMU data explained best the differences between the two types of conditions in the predictive model. Moreover, participants tended to move more during spontaneous smiles in the spontaneous block. In contrast, they moved more in between smiles during the posed block. This is unexpected, as the experimental setup heavily constrained the movements the participants could make. An explanation might be the different postures of the participants during the experiment. Further research should investigate other measures of head movement, and how they generalize to more ecologically valid setups.

Although the results achieved with the IMU data can be easily compared to those obtained with EDA, a direct comparison with the EMG results is not possible. This is mainly because of the different sampling rates of the sensors, and the frequency of the features used for classification. Further work is required in order to do a proper comparison of both.

A limitation of this study is that, due to the short stimuli, less smiles per participant were elicited. Only 27 participants self-tagged at least one spontaneous smile in the spontaneous block. In general, it is difficult to find a one-size-fits-all stimuli to make everybody smile or laugh. In our case, despite participants reported to smile more than 50% of the time, only about 66% of the participants showed any smile with the selected stimuli.

In this study no differences between Japanese, Chinese and Brazilians were observed regarding their smile temporal features. This seems to support the view of universal embodied responses. Despite the cultural differences of the participants involved in this study, there were no observed differences, even in posed smiles. One might expect that Japanese are better at posing smiles, given that their culture encourages frequent smiling to convey kindness and politeness. Even though the data

from the previous experiment seemed to point to that direction, this second study did not confirm the results. This might suggest that the ability to pose smiles is a personally trained one, and not learned by cultural context.

## 5.2 The multimodal signature of different smiles

Experiment 2 confirmed the potential of using EMG to identify the spatio-temporal dynamics of smiles, and their differences during different types of smiles (**RQ 2, 2-b**). Special care was taken in considering different types of posed smiles, including smiles posed for the camera, polite smiles, and smiles aimed at conveying the impression of having fun. Hence, these smiles seem to have similarities with other types of posed smiles, like the smiles posed under instruction.

By looking at the EMG signal, it was shown that the main differences between posed and spontaneous smiles lie in their temporal dynamics. This experiment further confirmed that rising time and decaying speed differ between both types of smiles (**RQ 2-b**).

Furthermore, it was found EDA and IMU measures alone have the potential to distinguish between co-occurring spontaneous and posed facial expressions. Previously, autonomic changes such as EDA have been found to co-occur with affective events [62], and those results are in line with the ones found here. Although no particular consensus has been reached on how EDA varies along other affective cues, the results of this experiment seem to suggest that EDA peaks anticipate facial expressions. Moreover, participants tended to move more during spontaneous smiles in the spontaneous block. In contrast, they moved more in between smiles during the posed block.

These results have to be considered carefully, as head movement artifacts might have caused the good results from the EDA measurements. In fact, IMU-measured data explained best the differences between posed and spontaneous events. This was already suggested by [40].

All in all, the results strongly support the potential robustness of a multimodal wearable approach to distinguish between spontaneous and posed positive affective responses (**RQ 3**).

No cultural differences found between posed and spontaneous smiles in this study (**RQ 5**). Furthermore, self-report measures also suggested that participants reacted and felt in a similar manner to the experimental tasks, regardless of their cultural background.

# Chapter 6

# The relationship between behavioral and electrophysiological responses and explicit self-report

As mentioned in Chapter 1 and Chapter 3, self-reported affect does not always correspond to the felt affect displayed by the wearer of the expression. Two main factors influence this mismatch. The first one is that the facial expression might be fabricated on purpose by the participants, regardless of their affective state. The second, is that participants might display a different amount of facial expressions than they are aware of. In this chapter, the correlation between smiling behavior and self-report is explored.

## 6.1 Electrophysiological activity and explicit self-report

The Affect Grid is a tool to report the valence and the arousal of a felt emotion. It is very convenient, because it allows to self-report the felt affective state in only one question.

Table 6.1: *Correlation between Affect Grid self-report and amount of spontaneous smiles.* The Pearson's correlation coefficients between the self-reported valence, arousal and the number of smiles coded by each coder. Significant correlations at $p < 0.01$ are marked with a *.

| | Valence | Arousal | Coder 1 | Coder 2 |
|---|---|---|---|---|
| Valence | 1.00 | 0.13 | **0.16*** | **0.16*** |
| Arousal | 0.13 | 1.00 | -0.05 | 0.06 |
| Coder 1 | **0.16*** | -0.05 | 1.00 | **0.45*** |
| Coder 2 | **0.16*** | 0.06 | **0.45*** | 1.00 |

In Experiment 1, participants reported their affective state after each stimuli. For some of the participants this was four times, for others, eight times. The self-reported data from all participants was gathered, and it was compared to the number of smiles that they displayed. The number of displayed smiles was labeled by two independent coders. Table 6.1 shows the Pearson's correlation coefficients between the self-reported valence, arousal and the amount of smiles coded by each coder.

As already suggested by the inter-coder agreement reported in section 4.1.4, the number of smiles identified by coder 1 and coder 2 are correlated. Moreover, the behavioral expressions are, albeit weakly, correlated with the self-reported valence. The moderate correlation might be because, in this experiment, participants were asked to conceal their facial expressions.

## 6.2 Behavioral and electrophysiological activity and the Affect Grid

Similarly to the previous section, the data from Experiment 2 was analyzed as well. In Experiment 2, participants self-reported their affective state in two occasions. Once after the posed and once after the spontaneous blocks. Table 6.2 shows the Pearson's correlation coefficients between the self-reported Valence and Arousal, and the number of smiles videocoded by each of the coders. Valence correlated with the number of smiles showed by each participant. However, this only happened with the amount of smiles coded by Coders 1 and 2, and not with the participant's own coding.

Table 6.2: *Correlation between Affect Grid self-report and amount of smiles.* The Pearson's correlation coefficients between the self-reported valence, arousal and the number of smiles coded by each coder. Significant correlations at p <0.01 are marked with a *. The participant's labels were correlated with those of Coder 2. Nevertheless, only Coder 1 and Coder 2 labels were correlated to valence scores.

| | Valence | Arousal | Participant's Coding | Coder 1 | Coder 2 |
|---|---|---|---|---|---|
| Valence | 1.00 | -0.07 | 0.13 | **-0.33*** | **-0.51*** |
| Arousal | -0.07 | 1.00 | 0.15 | 0.04 | 0.20 |
| Participant's Coding | 0.13 | 0.15 | 1.00 | **0.45*** | 0.20 |
| Coder 1 | **-0.33*** | 0.04 | **0.45*** | 1.00 | 0.71* |
| Coder 2 | **-0.51*** | 0.20 | 0.20 | **0.71*** | 1.00 |

On the other hand, arousal did not correlate with observable facial expressions. There are two possible reasons for this. First, participants did not report significant changes in arousal given the presented stimuli. Second, arousal might be related more to the intensity of the facial expressions, rather than to the existence of the expression itself.

Interestingly, the number of smiles coded by the participants correlated only with the labeling of Coder 1, who had knowledge of the experimental design.

## 6.3 Behavioral and electrophysiological activity and implicit self-report

In the previous two sections, a correlation between behavioral measures and explicit self-report was shown. Table 6.3 shows that both positive and negative affect reported implicitly, also correlates with the number of smiles coded by the participant and coder 1. It is interesting that in this measure, only the observations made by coders who had contextual information about the experimental design correlated with the measurement.

Table 6.3: *Correlation between IPANAT self-report and amount of smiles.* The Pearson's correlation coefficients between the self-reported IPANAT positive and negative scores, and the number of smiles coded by each coder. Significant correlations at $p < 0.01$ are marked with a *. The participant's coding, and the coder 1's coding was correlated to the IPANAT scores. Those were the ones that were aware of the experimental design.

| | Participant's Coding | Coder 1 | Coder 2 | IPANAT Positive | IPANAT Negative |
|---|---|---|---|---|---|
| Participant's Coding | 1.00 | **0.45*** | 0.20 | **0.23*** | **0.23*** |
| Coder 1 | **0.45*** | 1.00 | **0.71*** | **0.23*** | **0.23*** |
| Coder 2 | 0.20 | **0.71*** | 1.00 | 0.10 | 0.10 |
| IPANAT positive | **0.23*** | **0.23*** | 0.10 | 1.00 | 1.00* |
| IPANAT negative | **0.23*** | **0.23*** | 0.10 | **1.00*** | 1.00 |

## 6.4 Discussion

In this chapter it was shown that observed behavior was clearly related to the self-reported measures (**RQ 4**). This was confirmed in both experiments. In experiment 1, the correlation was weaker, as participants were asked to inhibit their facial expressions. In experiment 2, both coders' labels correlated to valence. The participants' did not.

Whereas explicit self-report correlated with the observations from both coders, only the self-reported video coding and coder 1's video coding correlated to IPANAT measures of affect. They were the only ones aware of the experimental design. This might suggest that the IPANAT measurement measures the contextual affective implications, rather than the ones felt by the participant (**RQ 4-b**).

# Chapter 7

# Human judgment of posed and spontaneous smiles

In previous chapters, the importance of measuring human experience was outlined, together with a proposal on how to automate such measurements. Furthermore, the potential of EMG to measure valence-related facial expressions was showed in the first two experiments. EMG turned out to be very good not only to measure fast and subtle spontaneous smiles, but also to distinguish between posed and spontaneous ones. The accuracy of EMG was calculated based on a ground truth composed of human coding of facial expressions, self-report, and most importantly, the experimental design used to collect the data. In section 3.5 and chapter 6, the challenges of establishing the ground truth were discussed. The facial expressions measurement can be done on a perceptual basis, considering only visible cues. However, a decision on whether the facial expression is posed or spontaneous is an inference. Table 5.1 showed that the perceptual judgment by independent coders yields a higher agreement than the inferential one. Moreover, the agreement between independent coders and the participant's is even lower. Therefore, electrophysiological and behavioral signals-based solutions seem to be more reliable than human judgments. However, the validity of these still depends on a certain degree on the human judgment used to establish the ground truth.

Since feelings and emotions are inherently private and specific to a person, finding a valid ground truth is of utmost importance. The difference between posed and spontaneous facial expressions is an inferential judgment, rather than a perceptual one. Thus, the ground truth should be based on good experimental design and self-report rather than on a third person's judgment. Experiment 3 was designed to further explore the degree in which a third person video coder would be able to distinguish between those two expressions, and agree with the experimental design and the self-reported label.

## 7.1 Experiment 3

### 7.1.1 Participants

73 voluntary participants took part on the study (37 female, average age=29 years old, SD=11). Participants were sampled from three nationality groups. 21 were Japanese (10 female), 20 were Chinese (10 female), and 32 were Mexican (17 female). The experiment was conducted in Japan, for Japanese and Chinese participants. Mexican participants were recruited and completed the task in Mexico.

### 7.1.2 Stimuli

54 smiles were selected from the smiles gathered in Experiment 2. These were 27 posed smiles and 27 spontaneous smiles. Each posed-spontaneous pair was produced by a participant of Experiment 2. In other words, the stimuli set contained videos of 27 participants (11 Brazilians, 9 Japanese, and 7 Chinese; 15 female), each of them smiling twice. Posed smiles were smiles happening in the posed block, and self-labeled as posed. Similarly, spontaneous smiles were smiles happening in the spontaneous block and self-labeled as spontaneous. Smiles of around 5 s were chosen to keep the experiment short. All stimuli had no sound to prevent contextual biases.

### 7.1.3 Experiment design and procedure

All participants went through all selected smiles. A computer program played the stimuli automatically. Participants were allowed to watch the stimuli only once. After watching the stimuli, two questions were presented one by one. These required them to report whether they considered the smile to be spontaneous or posed, and how confident they were about this judgment. After participants answered them, a continue screen showing how many videos they had watched, and how many they had left was shown on screen. Then participants were required to press the space bar once they were ready to continue to the next stimuli. After pressing continue, the next video started automatically, and the cycle was repeated. When participants had finished all videos, a screen showing their accuracy was presented. Accuracy was calculated by comparing the agreement between the participant's label, and the ground truth defined in experiment 2. Finally, a short interview was conducted to request demographic information and inquire about the strategy the participants used to make their judgments.

### 7.1.4 Measurements

Participants answered two questions per stimuli. The first was whether they though the smile was spontaneous or posed, with a forced choice between the two. The second was how confident they were about their judgment. A Visual Analogue Scale (VAS) was used to report confidence between 0 and 100%. At the end of the task, they answered demographic questions about their age, gender, and education level. Additionally, participants were asked what features did they use to make their judgments.

### 7.1.5 Apparatus

All stimuli were presented to the participant in a NEC Lavie Hz750/C laptop. The Python toolbox PsychoPy2 version 1.85.4 was used to create an automatic pre-

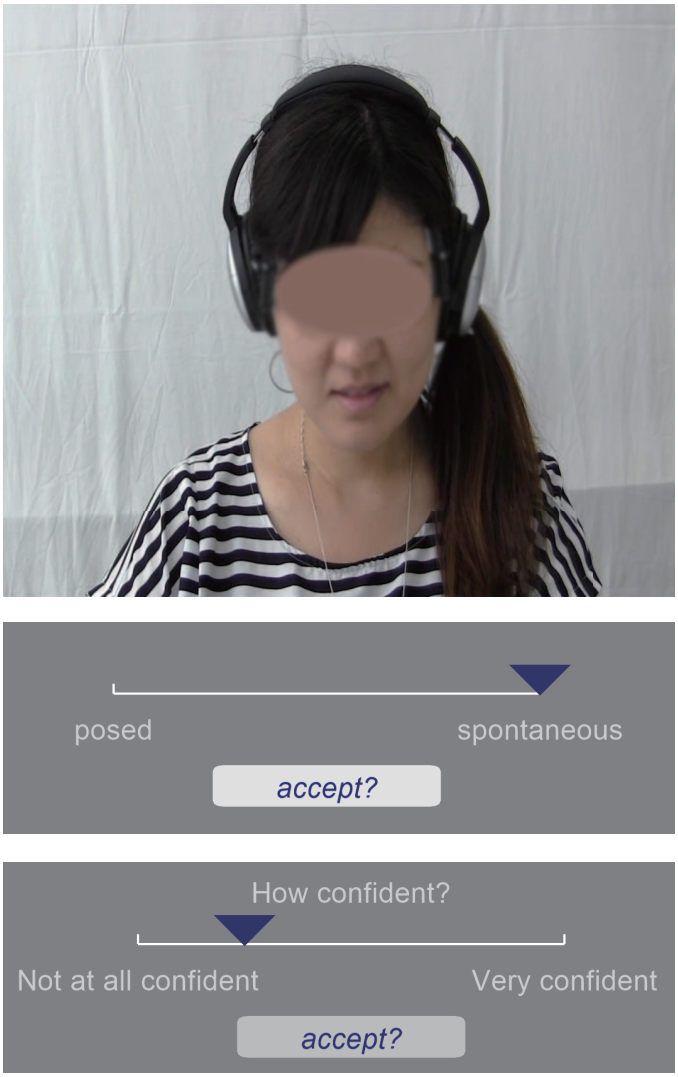



Figure 7.1: *Stimuli presentation and questions for experiment 3.* Participants had to watch the video of a smile lasting about 5 s, and then decide if it was posed or spontaneous. Afterwards, they reported how confident they were with their decision. An screenshot example showing the stimuli, and the post-stimuli questions are shown.

sentation of the stimuli, with the subsequent questions.

### 7.1.6 Results

The accuracy of every participant was calculated by comparing the label assigned by each participant and the ground truth established in Experiment 2. The overall accuracy of the participants in distinguishing posed and spontaneous smiles was

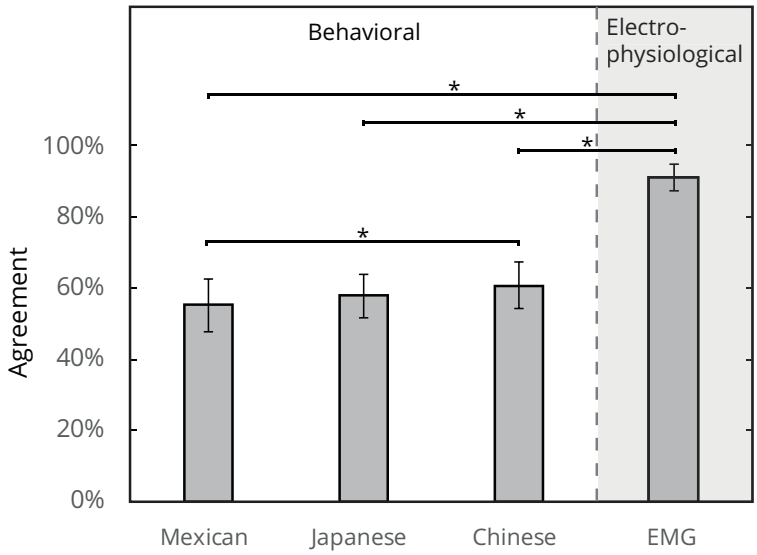



Figure 7.2: *Laypersons' accuracy when identifying posed and spontaneous smiles* Average accuracy scores are shown. There are significant differences between Chinese and Mexicans. Also, the difference between the machine's accuracy and the human groups is significant. The Y axis represents the accuracy in percentage. The X axis shows the groups.

0.58 (SD=0.073). A one sample t-test showed that the obtained accuracy is significantly different from a normal distribution with mean at chance level (mean=0.50, $t(72)=8.796$, $p < 0.01$). A one-way ANOVA with Nationality as between subjects independent variable showed a significant difference in the accuracy achieved by each group ($F(2,62)=3.754$, $p < 0.05$). Post hoc comparisons using the Tukey HSD test indicated that the mean score for Mexicans (M=0.55, SD=0.07) was significantly different from the score of Chinese nationals (M=0.61, SD=0.07). However, the Japanese nationals' score (M=0.58, SD=0.06) did not significantly differ from the Mexicans' and Chinese's.

Next, a one-way ANOVA comparing the achieved accuracy from the participants to the accuracy achieved by the spatio-temporal features classifier was performed. Significant differences were found between all nationalities and the machine learning performance ($F(3,96)=194.3$, $p < 0.001$). Figure 7.3 depicts a plot with average accuracies for all groups.

In average, all participants' confidence of their judgements was 0.71 (SD=0.16).

A 4-factor mixed design ANOVA with nationality and gender as between subjects factors; and smile category, and whether they guessed the correct label, as within subjects factors was performed on the reported confidence level. A significant difference was found between nationalities ($F(2,264)=17.453$, $p<0.001$) and gender groups ($F(1,264)=7.096$, $p<0.01$). The effect of smile category was not significant ($F(1,264)=0.383$, $p>0.05$). Similarly, their confidence level was not significantly dependent on whether they guessed correctly the ground truth label ($F(1,264)=0.038$, $p>0.05$). Finally, all other interaction effects were not significant. Post-hoc comparisons using the Tukey HSD test indicated that the mean score for Mexicans ($M=0.80$, $SD=0.14$) was significantly different ($p<0.001$) from the score of Chinese nationals ($M=0.66$, $SD=0.13$) and Japanese nationals ($M=0.61$, $SD=0.07$). However, the Japanese nationals' score did not significantly differ from the Chinese's.

Participants reported verbally on what features they based their decisions between posed and spontaneous smiles. Their responses were transcribed and analyzed using Affinity Diagram maps [124]. Figure 7.4(a) shows the results. Eight different features were found. The most participants (51 people, 70%) mentioned feature was eye movement, referring to the shape of the eyes, whether participants were gazing at the screen, or looking lost. Participants (34 people, 47%) also looked at body movements such as body vibration, shoulder shrinking, head movement, hiding the face, and the degree of relaxation of the posture of the person smiling. The next most popular feature (29 people, 40%) was mouth shape and movement. Participants looked at the opening and closing of the mouth, and whether the person smiling was showing the teeth or not. The next most common feature was the timing of the smile (18 people, 25%). This category grouped commentaries describing sudden changes, the duration of the smile, and the simultaneity of eyes and mouth movements. Afterwards, participants looked into the intensity of the smiles (13 people, 18%). By intensity, most participants mentioned how wide the person smiling was opening the mouth. Other less popular categories included eyebrow lifts (3 people, 4%), and surprisingly, the beauty of the smile (2 Chinese females, 3%). The percentage of people that

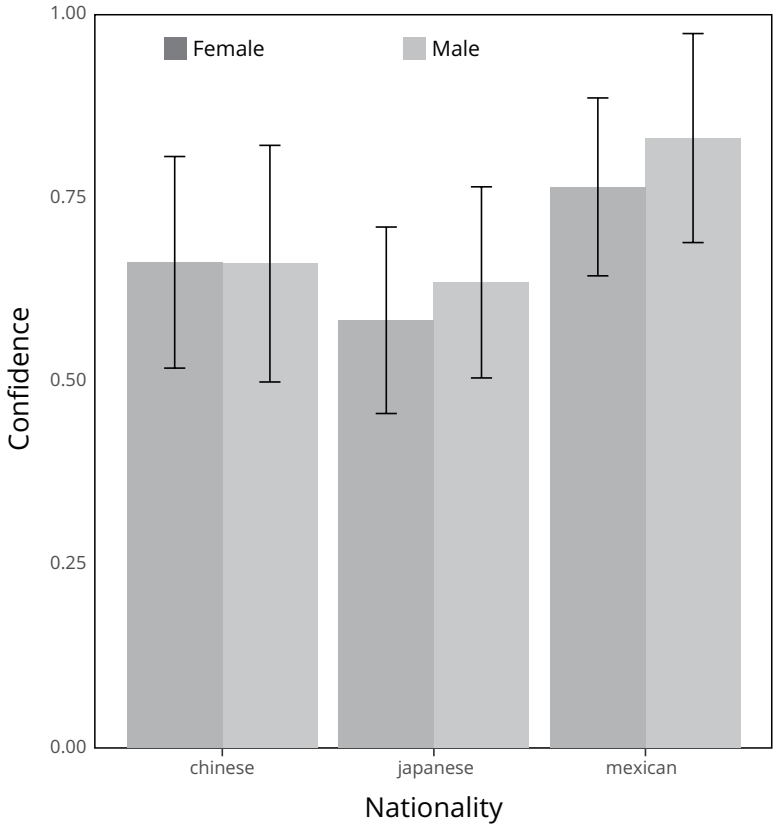
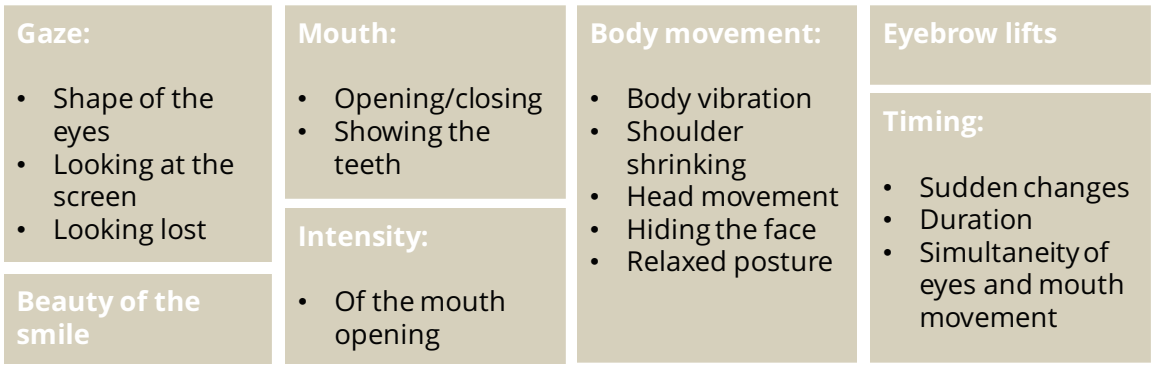



Figure 7.3: *Confidence on identification accuracy of posed and spontaneous smiles.* The self-reported confidence per nationality and gender is shown in the plot. According to the analyses described in section 7.1.6, only nationality and gender differences are significant. Mexicans were the most confident, even though their accuracy was the lowest.

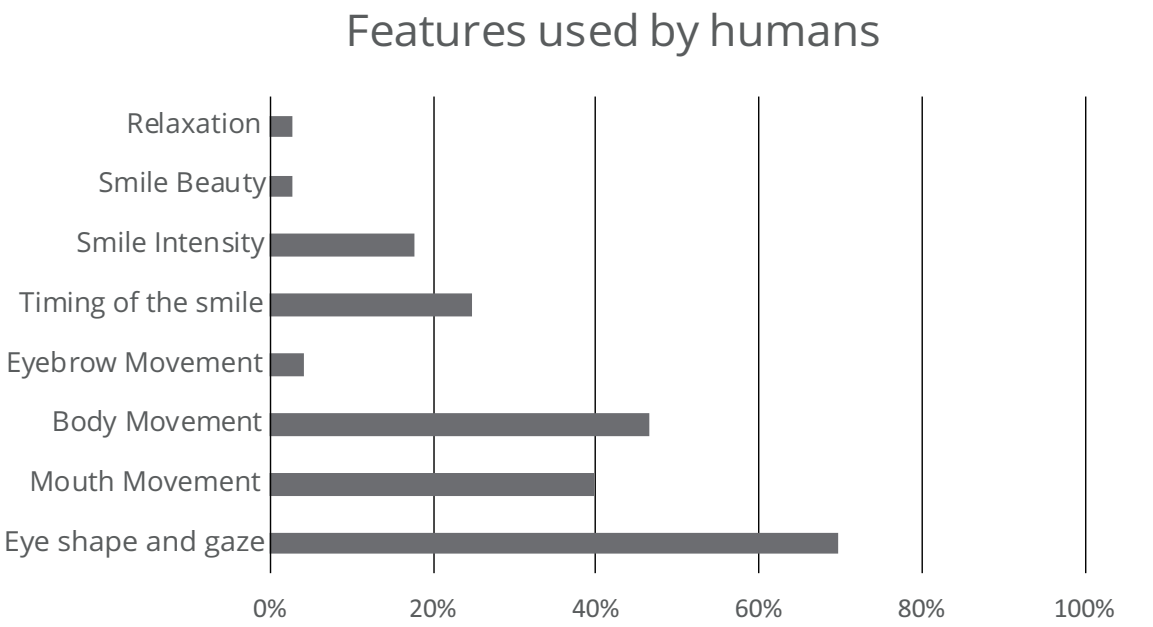
mentioned each feature is depicted in figure 7.4(b). It is important to note that the same person might have mentioned each feature more than once. Finally, an ANOVA was performed to test the influence of the features described on the task accuracy. From all the features, only the mouth feature had a significant effect on accuracy ($F(1,63)=7.9$, $p < 0.01$).

### 7.1.7 Discussion

To create an effective user experience automatic logging tool, identifying the ground truth with validity and reliability is of utmost importance. By using embodied cues of affect, this measurement can be done more effectively. However, the challenge




(a) *Affinity diagram.* The feature categories mentioned by the participants. Besides focusing on mouth and eye movement, intensity of the movement and speed, some people also looked at contextual cues as gaze; or subjective features such as beauty of the smile.




(b) *Feature usage percentage.* Human judges used eight different features to distinguish between posed and spontaneous smiles. The most common being eye movement, mouth movements, and body movements. Percentages show the proportion of people who mentioned each feature.

Figure 7.4: *Features used by laypersons to identify posed and spontaneous smiles.* The eight main features used by humans to distinguish between posed and spontaneous smiles are shown. The most common are eye, mouth, and body movements.

of establishing the ground truth remains. This is perhaps the most discussed topic among affective computing scientists and psychologists. In previous sections of this thesis, it was argued that perceptually established ground truths are only effective when discussing the existence of a visible embodied cue. However, more information

is required when high level judgments are made. This includes deciding whether a smile is posed or spontaneous. Results of experiment 2 showed that the agreement of independent coders differs greatly when doing this judgment. In this experiment, we further explored human performance when judging between posed and spontaneous smiles. Indeed, it was shown that relying only on visible behavioral cues is not as efficient as relying on electrophysiological cues. The results from the aforementioned algorithm were consistently better than a human's judgment based only on visible cues of the same data.

Even though the participant's accuracy scores were significantly higher than chance level, their accuracy was significantly lower than the machine's. Additionally, the 73 participants had different cultural backgrounds. The results suggested that Mexican's accuracy was significantly lower than Chinese's accuracy. Moreover, Japanese's accuracy was in between Mexican's and Chinese's, with no significant differences. This result might be related to the effect of practice in the task. The stimuli showed presented nationals from Japan, China, and Brazil. Japanese and Chinese were familiar with Asian faces, which are the majority on the sample. Furthermore, the Mexican sample was recruited in Mexico, where most of the participants had no or little experience dealing with foreigners. Even though Japanese people were recruited in Japan as well, their environment made it possible to get in touch with people from other nationalities. The other extreme is the Chinese sample. They were living in an environment where their native language was not spoken. Therefore, they would probably require paying more attention to non-verbal cues to understand conveyed messages. This might be an explanation why Chinese participants got the highest accuracy among the participants. An alternative explanation for the poor performance of Mexican participants is the Cross-Race Effect (CRE). This is a well-replicated finding in face recognition that people are better at recognizing faces from their own race, relative to other races[125, 126]. Since most of the people in the videos were Asians, Mexican's accuracy might have been decreased.

Interestingly, the self-reported confidence of Mexican participants was the high-

est, even though their performance was the lowest. This might be an instance of the Dunning–Kruger effect bias. This effect describes the fact that unskilled people not only "reach erroneous conclusions and make unfortunate choices, but their incompetence robs them of the metacognitive ability to realize it." [127]. Another factor that influenced confidence was gender. Males tend to be more confident than females, even though their accuracy do not differ.

According to the affinity diagram results, humans used eight features to distinguish between posed and spontaneous smiles. Eye, mouth, and body movement were the most common. Among all the features, mouth movement was the only reported feature that was significantly explaining the obtained accuracy. Furthermore, context seems to be really important for human judgement of different smiles. Many participants mentioned to have tried to guess the context of the people smiling. Specially where they were looking at. Several people mentioned that an idle gaze and smiling suddenly were posing.

All in all, in a context-less environment, EMG-based detection outperforms a layperson's judgment based on visible behavioral cues only. Even though the participants were not experts in reading posed and spontaneous smiles, they chose appropriate features to decide. From the eight features mentioned, the most common are relevant discrimination features according to the literature (see section 2.3). Nevertheless, it is difficult to distinguish posed and spontaneous smiles only from visible cues. Moreover, practice seems to improve the accuracy. The observed differences in the cultural background suggests that the practice gained by living abroad might have given an advantage to Chinese participants.

A limitation of this study is the reduced amount of information provided to the participants to make their choices. The stimuli provided were stripped from many contextual and multi-sensory cues that might have proven useful for the human participants. This was already shown in chapter 6, where the knowledge of the experimental design influenced the correlation between two independent coders and the participant. Since the specialty of humans is to integrate information from multimodal channels,

these results might not have favored them to take relevant decisions. It could be that they were too focused trying to guess the context, that they missed other important cues available through vision.

### 7.1.8 Conclusions

As shown in this experiment, humans can distinguish between posed and spontaneous smiles above chance level (**RQ 6**). However, their accuracy is very modest. Therefore, using behavioral and electrophysiological signals seems advantageous in this scenario. Using these signals is advantageous in this situation because these cues are not directly perceivable by humans. Finally, the reliability of these tools promises to complement the human ability to interpret information contextually (**RQ 6-b**).

# Chapter 8

# Discussion and potential applications

Emotions play an important role in our lives. They are often regarded as involuntary. Nevertheless, they drive most of the decisions we make, and they are powerful motivators. Hence, they are an important metric for evaluations of well-being and product quality. Emotions are assessed in different manners, including self-report and behavior coding. However, these tools have their drawbacks. The main of those being the subjective quality of emotion, and the biases that humans have when assessing them. Furthermore, they interrupt the user experience, and the low temporal resolution with which humans can be aware of their own feelings.

To assess quality of products and services, it is best when the true satisfaction of the user is identified. However, positive affective cues might be displayed out of politeness. A multimodal approach with spatio-temporal sensitive analysis is therefore proposed. The evaluation of this proposed method is made on positive affective responses, as they are a good measure of well being, and a prototypical example of cues that are intentionally expressed. Particularly, the focus is on distinguishing spontaneous and posed smiles.

In this thesis, a multimodal set of behavioral and electrophysiological measures is proposed to measure affective responses automatically and unobtrusively. Their

spatio-temporal dynamics are effective features in this task. Furthermore, identifying and counting affective responses is not enough to accurately assess human affect because these can be faked. Hence, extra judgments on whether these responses are felt or deliberately expressed are needed.

The proposed affective responses can be measured with a wearable device. The electrode positions are an advantage of the current device. They are optimized for smile detection without disrupting the movement of the face. Covering the face is an undesired situation as it makes the users artificially aware of their own facial expressions, and limits the movement of the skin [30,95]. In such cases, elicited facial expressions would be unnatural and definitely not spontaneous.

As first step, the feasibility of detecting spontaneous smiles using a surface EMG wearable was evaluated. Even when the smiles are at the micro-expression level. The results showed that micro-smiles can be distinguished from a neutral face using EMG, with good accuracy. Therefore, EMG pre-processing and classification methods seem to be also useful to analyze micro-smiles. Using EMG for this purpose has several advantages. The temporal resolution and portability of this device would allow to provide real-time feedback, if desired. This can be used for quantification applications of positive facial expressions. A limitation is the need for individual calibration to ensure good performance. Moreover, eliciting spontaneous expressions for calibration is a challenging task.

The main limitation of the first study was that smiles tend to last longer than half a second [117]. They reported that expressions of happiness tend to last longer than the micro-expression threshold. When the stimuli are strong enough, the displayed affective expression is often long-lasting laughter. It requires quite some effort to conceal laughter. While some participants are very good in neutralizing their facial expressions, most others are not. Therefore, the number and nature of the spontaneous smiles elicited was very specific to the elicitation method.

In chapter 4, the first effort to prove the feasibility of detecting micro-smiles with a wearable device was described. This is an important first step for automatic analysis

of spontaneous smiles and laughter in human-human communication. As observed from our results, in ecologically valid settings people tend to accompany laughter with head and hand movements. Therefore, other major expressive modalities such as speech, body movements, and postural attitudes, might be complementary to annotate the situation and eventually infer its meaning.

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

This is specially advantageous in the task of distinguishing posed from spontaneous smiles. Further data analysis on experiment 1, provided evidence on the EMG's ability to measure features unavailable to the human eye, and vision in general. Good accuracy results were achieved in the task of distinguishing between posed and spontaneous smiles. The results described in section 4.2.2 show that the algorithm using spatial and magnitude EMG features achieved accuracy rates with more variability than the algorithm using the spatio-temporal features. This suggests that even though some participants can fake how a smile should look like to appear spontaneous, they cannot mimic its temporal profile. Furthermore, the proposed algorithms take advantage of the ICA extraction to estimate different sources of the EMG signal and its magnitude. They also profit from the EMG's high temporal resolution to estimate smile characteristics without consuming excessive computational resources. From these two alternatives, the most successful results were given by considering the temporal resolution of the signal. As supported by previous studies, this is probably because spontaneous smiles and posed smiles differ in this aspect. In this data, the

main difference was that spontaneous smiles tend to last longer than posed ones. Moreover, the spatio-temporal scores of Japanese nationals vary less than those of other nationalities. This suggested that in a more balanced sample, cultural differences might be observable.

Despite the good results achieved, there are some limitations in this study. First of all, spontaneous smiles were elicited using a micro-expression elicitation paradigm. This implies that their dynamics might differ from other smiles. The unavailability of micro-smile examples causes a significant unbalancing in the data. This, in turn, might bias the performance of the machine learning algorithm. Another possible confounding is that the duration of the posed smiles might be constrained by the duration of the instruction used to elicit them. Experiment 2 addressed these limitations to obtain high-quality data.

In the second experiment (section 5.1), posed and spontaneous smiles were elicited and analyzed. Spontaneous smiles were elicited by showing positive-valenced videos to the participants. Posed smiles were asked, even during an slightly unpleasant situation. This experimental design allowed to some extent control the affect felt by the participants when they produced the required smiles. As a validation check, self-reported measures and self-video coding were asked. Self-reported measures were in line with the experimental design. Participants reported feeling more positive during the spontaneous block than during the posed block, independently from their nationality or gender. On the other hand, no arousal differences were reported between spontaneous and posed blocks. Similar to the arousal, the reported IPANAT scores presented no differences between experimental blocks. In this case, Japanese seemed to give higher scores for both positive and negative affect than the Brazilians. This might be because in Japanese language, many contextual inferences are made. Furthermore, the correlations between the video-coded data from people with contextual knowledge and the IPANAT scores suggested the importance of context awareness in the IPANAT scores. Furthermore, the lack of correlation between explicit self-report (Affect Grid) and Implicit self-report (IPANAT) suggest that both measurement tools

are measuring different levels of affective awareness. Moreover, explicit self-report is subject to the demand characteristics bias. This describes the tendency of participants to play the role of good subjects and respond according to their guess of what they are expected to answer [125, 128].

Both experiment one (section 4.2.3) and two (section 5.1.7) showed that the duration of these smiles did differ. The duration difference might have been influenced by the elicitation tasks in Experiment 1. Hence, Experiment 2 took special care to elicit posed smiles that deliberately intend to convey the impression of having fun. Despite the different elicitation tasks, rising time was still significantly different between the two types of smiles in both experiments. Furthermore, decaying speed turned out to significantly differ. As in the previous experiment, magnitude of posed and spontaneous smiles also did not differ. High accuracy was maintained by using spatio-temporal features in the second data set. This corroborates the robustness of the proposed algorithm, for smiles elicited with different paradigms. Future work should explore the differences between various posed smiles. In experiment 1, the posed smiles were smiles for the camera, whereas in experiment 2, the smiles tried to convey the message of having fun. Other types of smile might also exist. For example, a polite smile. During experiment 2, it was also observed that participants smiled politely to the experimenter when the experiment was over. In future work, it would be interesting to also explore this type of smile.

It is important to notice that the differences found by the algorithm cannot fully explain why those differences are caused. In an attempt to further explore this, filtered EMG ratios were analyzed. However, the differences found did not show any specific trend. The EMG activity of the participants differs among experimental blocks, nationality, and channel. It seems that the EMG activation from Chinese participants has a wider Standard Deviation than the other two nationality groups. Furthermore, the activity of the right-sided channels, is higher than those placed on the left side. A possible explanation is the expected symmetry of an enjoyment smile, as opposed to the asymmetry of a fake smile [37]. On the other hand, this might

have been measurement noise, as the effect disappears after the ICA Blind-Source Separation (BSS) algorithm is applied. The magnitudes of the temporal features per IC do not significantly differ. Only the interaction between rising speed and IC. These results highlight the advantage of the wearable system used. Even though the EMG is measured distally, the Blind Source Sepparation performed by the ICA allows the system to accurately estimate the relevant components caused by EMG activity. Even though this system does not allow to measure the exact source of the muscle activity, it allows to record such activity unobtrusively and yet accurately. Even though the EMG is measured distally, the activity of different muscular sources can be estimated and used for feature extraction.

Experiment 2 allowed to explore the role of autonomic responses during both posed and spontaneous smiles. Skin conductance on the hands changes more prominently than from the neck. However, the both measures are correlated. The same algorithm was used to process both signal sources. An effort was made to reduce motion artifacts, and different features were calculated to train a SVM to classify between posed and spontaneous events. The results of this algorithm were very good. However, commonly used features such as magnitude, first order derivatives, and peaks per smile have the disadvantage that only one feature can be calculated per smile. Hence, magnitude of the EDA alone was preferred to increase the data available for training and testing. The success of the magnitude features alone compared to other features might be due to the amount of data available.

Both the magnitude differences between neck and hand EDA were proven to be suitable for the task at hand. However, the differences in their magnitudes were only marginally significant. Form visual inspection of the measured EDA responses, it seems that EDA changes anticipate facial movement [62]. Further analyses are needed to confirm this hypothesis.

Moreover, the obtained EDA results have to be considered carefully. Although the pre-processing was chosen to reduce motion artifacts, these artifacts might explain the high accuracy obtained. Using motion sensing yields higher accuracy than that

of EDA. This suggests that motion can explain the differences between spontaneous and posed affective responses quite well. Thus, the covariance of EDA and IMU data might explain the predictive power of EDA, even if participants did not explicitly report to be aroused. Furthermore, the differences in magnitudes between blocks were non significant. The head IMU data explained best the differences between the two types of conditions. This is unexpected, as the experimental setup heavily constrained the movements the participants could make. A possible reason is that participants seated in different postures during both conditions. Further research should investigate whether these results generalize to more ecologically valid setups.

Although the results achieved with the IMU data can be easily compared to those obtained with EDA, a direct comparison with the EMG results is not possible. This is mainly because of the different sampling rates of the sensors, and the frequency of the features used for classification. Further work is required in order to do a proper comparison of both.

In this study, no differences between Japanese, Chinese and Brazilians were observed regarding the temporal features on their smile's EMG. This seems to support the view of universal embodied responses. Despite the cultural differences of the participants involved in this study, there were no observed differences, even in posed smiles. One might expect that Japanese, Chinese and Brazilians learn to be polite in a different manner, or that they smile spontaneously with different frequency. Even though the data from experiment 1 seemed to point to that direction, this second study did not confirm the results. This in turn, seems to support the view of universal embodied responses. The ability to pose smiles might therefore be a personally trained one, and not learned by cultural context.

In chapter 6, the relationship between self-reported affect and smiling behavior was reported. Self-report and behavioral cues such as facial expressions do not always correspond to the felt affect, or to each other. It was argued that two main factors influence this mismatch. The first one is that the facial expression might be fabricated on purpose by the participants, regardless of their affective state. The sec-

ond, is that participants might display a different amount of facial expressions than they are aware of. The results showed that observed behavior was clearly related to the self-reported measures. Even when participants were trying to conceal their facial expressions, these were significantly correlated to self-reported affect. During experiment 2, explicitly reported affect correlated with both coders, but not with the coding of the participants themselves.

On the other hand, implicitly reported affect correlated with the video-coding of the coders who had contextual information about the experiment design. Furthermore, they did not correlate to each other. This suggests that both measures of affect measure different levels of awareness. It might seem that the explicit measure is related more on how aware the participant was of the purpose of the experiment, and affected by demand characteristics. This might also have affected the implicit report to a certain extent, as it was requested after the explicit self-report. Thus, participants were, most probably, already aware of the purpose of the questionnaire. Further work should explore the differences between both types of measures. It would also be interesting to explore how the level of awareness of the participant relates to both behavior and self-report.

To create an effective user experience automatic logging tool, identifying the ground truth with validity and reliability is of utmost importance. By using embodied cues of affect, this measurement can be done more effectively. However, the challenge of establishing the ground truth still remains. Perceptually established ground truths are only effective when discussing the existence of a visible behavioral cue. However, more information is required when high level judgments are made. This includes deciding whether a smile is posed or spontaneous. Results of experiment 2 showed that the agreement of independent coders differs greatly when doing this inferential judgment (table 5.1). Experiment 3 further explored human performance when judging between posed and spontaneous smiles. It was shown that a judgment based on electrophysiological cues outperforms a layperson's judgment based on behavioral cues alone. The results from the aforementioned algorithm were consistently better

than a human's judgment of the same data. Therefore, the traditionally used "third person coder labeling" would not be enough in this case.

Participants with different cultural backgrounds achieved different identification performance averages. The results suggested that Mexican's accuracy was significantly lower than Chinese's accuracy. Japanese's accuracy was in between the other two. This result might be related to the effect of practice in the task. Practice also relates to the contextual opportunity and importance to use behavioral information to convey a message that the participants had during their everyday life. The Mexican sample had no or little experience dealing with foreigners. The other extreme is the Chinese sample. They were living in an environment where their native language was not spoken and rely heavily on body language. This might be an explanation why Chinese participants got the highest accuracy among the participants. Interestingly, the self-reported confidence of Mexican participants was the highest, even though their performance was the lowest. This might be an instance of the Dunning–Kruger effect bias. Moreover, the Cross-Race Effect might have hindered the Mexican's recognition of Asian's facial expressions. Since two thirds of the participants in Experiment 2 were Chinese and Japanese, most of these did not match their own race.

According to the affinity diagram results, humans used eight features to distinguish between posed and spontaneous smiles. Eye, mouth, and body movement were the most common. Among all the features, mouth movement was the only reported feature that was significantly explaining the obtained accuracy. Furthermore, context seems to be of utmost importance for human judgment of different smiles. Several participants mentioned to have tried to guess the context of the people smiling.

All in all, in a context-less environment, sensor-based detection outperforms a layperson's inferential judgment. Even though the participants were not experts in reading posed and spontaneous smiles, they chose appropriate features to decide. The observed differences in the cultural background suggests that the practice gained by living abroad might have given an advantage to Chinese participants.

A limitation of the third experiment is the reduced amount of information provided

to the participants to make their choices. The stimuli provided were stripped from many contextual and multi-sensory cues that might have proven useful for the human participants. This was already shown in previous chapters, where the knowledge of the experimental design influenced the correlation between two independent coders and the participant. Since the specialty of humans is to integrate information from multimodal channels, these results might not have favored them to take relevant decisions. It could be that they were too focused trying to guess the context, that they missed other important available cues.

Finally, it is worth mentioning that besides the performance in accuracy, using automatic judgments is also more practical than asking a person to label the data. In experiment 2, participants took around 30 minutes to judge 54 videos, each of about 10 s or less. The algorithm is able to do the judgment of the same amount of data in seconds. Thus, not only can a wearable solution perceive the most relevant information, it can also process the acquired data more efficiently.

## 8.1 Potential applications

The results of this study showed the potential of using a wearable system to measure positive affective cues in a valid and reliable manner. Such system has several potential applications for research and user evaluation of products and services. Some of the applications include:

- **A multimodal wearable device for quantification of affective responses in iterative design processes.** This system would allow to assess experience sampling without experience disruption. Since the quantified behavior is correlated to the self-reported measures, asking would not longer be necessary.
- **Scene understanding of affective salient elements on a scene or frames in a video.** This would be useful to assess the effectiveness of media, specially advertisement. Frames where smiles occur can be deemed as the most interesting.

- **Predicting user judgment of media content.** With this system, it would be possible to assess whether a viewer is really enjoying a media product or not. With this information, successive media recommendations could be possible.
- **Estimating self-reported affective states for psychological and user research.** The unobtrusiveness of the wearable makes it a good alternative to measure fast psychological reactions to a multitude of stimuli and social interactions.

The high accuracy and high temporal resolution makes of this system an excellent system for research, specially in environments with high degree of movement or occlusion. It also provides an advantage where two or more persons have to be measured simultaneously. In these contexts, other measurement techniques would be time consuming to implement, too computationally complex, or inaccurate.

# Chapter 9

# Conclusions and future directions

Quantification of affective responses is useful for different applications. For some of these, counting affective cues is enough, and the interpretation is left for a field expert. However, these cues are prone to biases and manipulations from the person displaying or reporting the emotion. Therefore, methods to overcome such biases have to be considered.

This is a challenging task. Perhaps the most important challenge is to establish a ground truth. Specially for perceptual judgments, the best approximation of the ground truth is a combination of good experimental design, observation, and self-report. Based on that ground truth, a wearable system using both electrophysiological and behavioral cues could arguably outperform the judgment of a human coder who relies only on visual cues. Therefore, training machines to do this type of judgment would be beneficial.

Along three experiments, a multimodal wearable system was proposed to assess the dynamics of different positive affective responses. Both behavioral and electrophysiological signals were chosen. Particular interest was taken in distinguish the differences between affective responses during both posed and spontaneous events. Particularly, six research questions were asked and answered.

**RQ1. Is it feasible to use EMG to detect positive fast and subtle facial expressions at the micro-expression level?** Using the data collected from experiment 1, it was shown that distal EMG is an effective measure to identify fast and subtle spontaneous smiles, even at a micro-expression level. This method is specially advantageous when two or more people are being tracked (**RQ 1-b**), as occlusion and movement is common in this situation.

**RQ2. Can we distinguish a positive affective posed reaction from a spontaneous one using EMG?** It is possible to identify the differences between both types of smiles using EMG. Whilst the spatial distribution of the muscles used to make these smiles differ, it was found that temporal features are more robust to distinguish between posed and spontaneous smiles (**RQ 2-b**). Namely, smile duration, rising time, and decaying time.

A second experiment confirmed the potential of using EMG to identify the spatio-temporal dynamics of smiles, and their differences. Different types of posed smiles were considered. In this case, rising time and decaying speed differed significantly (**RQ 2-b**).

**RQ3. How does a multimodal system, including head movement and EDA, improve the identification of posed and spontaneous positive affective responses?** EDA and IMU measures alone have the potential to distinguish between co-occurring spontaneous and posed facial expressions. The results seem to suggest that EDA peaks anticipate facial expressions. Moreover, IMU-measured data explained best the differences between posed and spontaneous events, as people tend to move more during spontaneous smiles.

**RQ4. How does the observed behavior occurrence and dynamics relate to self-reported measures of affect?** Observed behavior was clearly related to the self-reported measures. This was confirmed in both experiments. When participants were asked to inhibit their facial expressions, the correlation was weaker. Whereas

explicit valence self-report correlated with the observations from both coders; only the video coding of people aware of the experimental design correlated with implicit self-report. This might suggest that both measurements relate to different affective awareness levels (**RQ 4-b**).

**RQ5. How does cultural background affect our affective responses regarding the investigated measures?** No cultural differences found between posed and spontaneous smiles in this study. Therefore, the results support the view that production of embodied affective responses are similar for all humans, in spite of their learned cultural background.

**RQ6. How good are humans at distinguishing between posed and spontaneous smiles?** Humans can distinguish between posed and spontaneous smiles above chance level. However, their accuracy is very modest. Using behavioral and electrophysiological signals seems advantageous in this scenario. These cues are not directly perceivable by humans, therefore they provide complementary information. Finally, the reliability of these tools promises to improve an expert's ability to interpret information contextually (**RQ 6-b**).

## 9.1 Contributions to the Human Informatics field

The contributions of this work to the multidisciplinary field of Human Informatics are outlined below. These contributions lie in the intersection between engineering, psychology, and design.

- A multimodal wearable approach was proposed and successfully implemented to assess different types of affective responses.

- It is possible to detect fast and subtle spontaneous smiles from distal EMG. This is advantageous in multiple-users settings.

- The high temporal resolution of EMG-based detection is advantageous in the task of identifying different facial expressions, and particularly, their temporal dynamics.

- The temporal dynamics of the EMG of a smile to distinguish posed and spontaneous smiles were analyzed. Spontaneous smile duration differs from posed smile duration; magnitude both types of smiles is not significantly different; and rising time, and decay time differ.

- Two algorithms to distinguish posed and spontaneous smiles from EMG were developed. Distinguishing between posed and spontaneous smiles using EMG is moderately successful when using spatial and magnitude features. Performance increases when temporal features are included. Moreover, spatio-temporal features seem more robust for individual differences.

- No cultural differences were observed between in the temporal dynamics of different smiles.

- The developed system outperformed humans using only visual cues in the task of distinguishing between posed and spontaneous smiles.

- A method was proposed to elicit posed and spontaneous smiles in a controlled setting, and taking care of obtaining a balanced amount of expressions.

- Looking at embodied cues of affect such as facial expressions can be a valid alternative to self-report. When these cues are measured automatically, the reliability of the measurements is also increased.

- This wearable approach might be arguably better than other automatic identification technologies, as the main difference between posed and spontaneous smiles lies on their temporal dynamics.

- The ground truth for such automatic identification systems must not be a third person's judgement, as this seldom agrees with the ground truth established by

the elicitation method and self-report.

- Quantifying electrophysiological and behavioral cues in an automatic fashion will help to accelerate research on the area of human informatics. With this methodology is feasible to continuously measure human experience in a reliable, and comfortable manner. Access to this data will support psychologists, designers, marketeers and therapists to study these human information to their advantage.

## 9.2 Future Work

The present work outlines the continuation of years of psychological studies. Even though many have asked the same questions, to the best of the author's knowledge, none has addressed them with a distal facial EMG-based wearable approach. Second, it is the first time that identification of posed and spontaneous smiles from co-occurring EDA was studied. Finally, head movement was confirmed as an important predictor of spontaneity. Additionally, it was argued that identification of smiles and other facial expressions should not be limited to perceptual detection, but inferential judgments made by automatic systems would also be helpful in a number of applications.

This is a first step to readdress ancient questions with new wearable technology. The main advantage of the proposed method is the wearability of the device, and the possibility of measuring EMG distally. I hope that this work will inspire many researchers to keep investigating the embodied nature of our emotions.

Future work should keep considering carefully the validity of the chosen ground truth. This wearable approach will be a tool to explore with more temporal resolution how emotion processes arise and develop in our bodies, and explore embodied affective responses in the wild.

Moreover, this work has only proven the potential of the proposed approach. Further work should consider developing a between-subjects model, real-time imple-

mentation, and improving the wearability of the proposed system. This includes the technical challenges of making it robust against noise, and with long-lasting batteries. This would allow its usage in more ecologically valid scenarios.

# Bibliography


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

[2] A. Thieme, J. Wallace, T. D. Meyer, and P. Olivier, "Designing for Mental Wellbeing : Towards a More Holistic Approach in the Treatment and Prevention of Mental Illness," in *Proceedings of the 2015 British HCI Conference*. ACM, 2015, pp. 1–10.

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

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

[5] M. Soleymani, S. Asghari-esfeden, Y. Fu, and M. Pantic, "Analysis of EEG Signals and Facial Expressions for Continuous Emotion Detection," *IEEE Transactions on Affective Computing*, vol. 7, no. 1, pp. 17–28, 2016.

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

[7] J.-Y. Mao, K. Vredenburg, P. W. Smith, and T. Carey, "The state of user-centered design practice," *Communications of the ACM*, vol. 48, no. 3, pp. 105–109, 2005.

[8] J. P. Jokinen, "Emotional user experience: Traits, events, and states," *International Journal of Human Computer Studies*, vol. 76, pp. 67–77, 2015.

[9] M. Csikszentmihalyi and R. Larson, "Validity and Reliability of the Experience-Sampling Method," in *Flow and the Foundations of Positive Psychology*. Dordrecht: Springer Netherlands, 2014, pp. 35–54.

[10] R. Larson and M. Csikszentmihalyi, "The Experience Sampling Method," *New Directions for Methodology of Social & Behavioral Science*, vol. 15, pp. 41–56, 1983.

[11] A. J. Nederhof, "Methods of coping with social desirability bias: A review," *European Journal of Social Psychology*, vol. 15, no. 3, pp. 263–280, jul 1985.

[12] J. G. Adair, "The Hawthorne effect: A reconsideration of the methodological artifact." *Journal of Applied Psychology*, vol. 69, no. 2, pp. 334–345, 1984.

[13] J. A. Russell, A. Weiss, and G. A. Mendelsohn, "Affect Grid: A Single-Item Scale of Pleasure and Arousal," *Journal of Personality and Social Psychology*, vol. 57, no. 3, pp. 493–502, 1989.

[14] M. Bradley and P. J. Lang, "Measuring Emotion: The Self-Assessment Semantic Differential Manikin and the," *Journal of Behavior Therapy and Experimental Psychiatry*, vol. 25, no. I, pp. 49–59, 1994.

[15] A. Betella, P. F. M. J. Verschure, B. Cuthbert, S. Viger, N. Novak, and A. Berger, "The Affective Slider: A Digital Self-Assessment Scale for the Measurement of Human Emotions," *PLOS ONE*, vol. 11, no. 2, p. e0148037, feb 2016.

[16] M. M. van der Ploeg, J. F. Brosschot, J. F. Thayer, and B. Verkuil, "The Implicit Positive and Negative Affect Test: Validity and Relationship with Cardiovascular Stress-Responses." *Frontiers in psychology*, vol. 7, p. 425, 2016.

[17] B. K. Payne, C. M. Cheng, O. Govorun, and B. D. Stewart, "An inkblot for attitudes: Affect misattribution as implicit measurement." *Journal of Personality and Social Psychology*, vol. 89, no. 3, pp. 277–293, sep 2005.

[18] F. Qu, W.-J. Yan, Y.-H. Chen, K. Li, H. Zhang, and X. Fu, ""You Should Have Seen the Look on Your Face...": Self-awareness of Facial Expressions," *Frontiers in Psychology*, vol. 8, p. 832, may 2017.

[19] K. R. Scherer, "Unconscious processes in emotion," in *Emotion and Consciousness*, L. Barret, P. Niedenthal, and P. Winkielman, Eds. New York: Guilford Press, 2005, ch. 13, pp. 312–344.

[20] J. H. Janssen, P. Tacken, J. G.-J. de Vries, E. L. van den Broek, J. H. Westerink, P. Haselager, and W. a. IJsselsteijn, "Machines Outperform Laypersons in Recognizing Emotions Elicited by Autobiographical Recollection," *Human–Computer Interaction*, vol. 28, no. 6, pp. 479–517, 2013.

[21] A. Vinciarelli, M. Pantic, and H. Bourlard, "Social signal processing: Survey of an emerging domain," *Image and Vision Computing*, vol. 27, no. 12, pp. 1743–1759, 2009.

[22] R. A. Calvo, S. Member, S. D. Mello, and I. C. Society, "Affect Detection : An Interdisciplinary Review of Models , Methods , and Their Applications," *IEEE Transactions on Affective Computing*, vol. 1, no. September, pp. 18–37, 2010.

[23] P. Ekman and W. P. Friesen, "Measuring facial movement with the Facial Action Coding System," in *Emotion in the human face*, 2nd ed., P. Ekman, Ed. Cambridge University Press, 1982, ch. 9, pp. 178–211.

[24] J. A. Russell, "A circumplex model of affect," *Journal of Personality and Social Psychology*, vol. 39, no. 6, pp. 1161–1178, 1980.

[25] K. R. Scherer, A. Schorr, and T. Johnstone, *Appraisal processes in emotion : theory, methods, research.* Oxford University Press, 2001.

[26] P. Ekman, "Basic Emotions," in *Handbook of cognition and emotion*, T. Dalgleish and M. Power, Eds. John Wiley & Sons, Ltd., 1999, ch. 3, pp. 45–60.

[27] J. A. Russell, "Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies." *Psychological Bulletin*, vol. 115, no. 1, pp. 102–141, 1994.

[28] V. Bettadapura, "Face Expression Recognition and Analysis: The State of the Art," *CoRR*, pp. 1–27, 2012.

[29] Y. Chen, Z. Yang, and J. Wang, "Eyebrow emotional expression recognition using surface EMG signals," *Neurocomputing*, vol. 168, pp. 871–879, 2015.

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

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

[32] P. Ekman, "Darwin, Deception, and Facial Expression," *Annals of the New York Academy of Sciences*, vol. 1000, pp. 205–221, 2003.

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

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

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

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

[37] P. Ekman and E. Rosenberg, *What the face reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS)*, 2nd ed., P. Ekman and E. Rosenberg, Eds. Oxford University Press, 2005.

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

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

[40] M. F. Valstar, H. Gunes, and M. Pantic, "How to distinguish posed from spontaneous smiles using geometric features," *International Conf. Multimodal Interfaces*, vol. terfaces, pp. 38–45, 2007.

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

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

[43] M. F. Valstar, H. Gunes, and M. Pantic, "How to distinguish posed from spontaneous smiles using geometric features," *Int'l Conf. Multimodal In*, vol. terfaces, pp. 38–45, 2007.

[44] F. Silveira, B. Eriksson, A. Sheth, and A. Sheppard, "Predicting audience responses to movie content from electro-dermal activity signals," in *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing - UbiComp '13*. New York, New York, USA: ACM Press, 2013, p. 707.

[45] P. Ekkekakis, "Mood, Emotion and affection," in *Measurement in sport and exercise psychology*, G. Tenenbaum, R. Eklund, and A. Kamata, Eds. Human Kinetics, 2015, ch. 28, pp. 321–332.

[46] C. Darwin, *The Expression of the Emotions in Man and Animals*. New York D. Appleton and Company, 1872.

[47] A. Öhman, "Face the Beast and Fear the Face: Animal and Social Fears as Prototypes for Evolutionary Analyses of Emotion," *Psychophysiology*, vol. 23, no. 2, pp. 123–145, mar 1986.

[48] W. James, "What is an emotion?" *Mind*, vol. os-IX, no. 34, pp. 188–205, apr 1884.

[49] A. R. Damasio, *Descartes' error : emotion, reason, and the human brain*. Putnam, 1994.

[50] A. Damasio, "The somatic marker hypothesis and the possible functions of the prefrontal cortex," *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, vol. 351, no. 1346, pp. 1413–1420, Oct 1996.

[51] R. B. Zajonc, "Feeling and thinking: Preferences need no inferences." *American Psychologist*, vol. 35, no. 2, pp. 151–175, 1980.

[52] W. B. Cannon, "The James-Lange Theory of Emotions: A Critical Examination and an Alternative Theory," *The American Journal of Psychology*, vol. 100, no. 3/4, p. 567, 1987.

[53] M. B. Arnold, *Emotion and Personality. Psychological Aspects*. New York: Columbia University Press, 1960.

[54] R. Lazarus, *Emotion and adaptation*. Oxford, UK: Oxford University Press, 1991.

[55] P. Ekman, "An Argument for Basic Emotions," *Cognition and Emotion*, vol. 6, no. 3-4, pp. 169–200, 1992.

[56] J. R. Fontaine, K. R. Scherer, E. B. Roesch, and P. C. Ellsworth, "The World of Emotions is not Two-Dimensional," *Psychological Science*, vol. 18, no. 12, pp. 1050–1057, dec 2007.

[57] J. J. Prinz, *Gut reactions : a perceptual theory of emotion*. Oxford University Press, 2004.

[58] D. Galati, K. R. Scherer, and P. E. Ricci-Bitti, "Voluntary facial expression of emotion: comparing congenitally blind with normally sighted encoders." *Journal of personality and social psychology*, vol. 73, no. 6, pp. 1363–79, dec 1997.

[59] P. Ekman and E. L. Rosenberg, *What the Face RevealsBasic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS)*, 2nd ed. Oxford University Press, apr 2005.

[60] R. Buck, "Nonverbal behavior and the theory of emotion: the facial feedback hypothesis." *Journal of personality and social psychology*, vol. 38, no. 5, pp. 811–24, may 1980.

[61] S. D. Kreibig, "Autonomic nervous system activity in emotion: A review," *Biological Psychology*, vol. 84, no. 3, pp. 394–421, jul 2010.

[62] W. Boucsein, *Electrodermal Activity*. Boston, MA: Springer US, 2012.

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

[64] W. Ruch and P. Ekman, "The expressive pattern of laughter," in *Emotion qualia, and conciousness*, A. Kaszniak, Ed. Tokyo: World Scientific Publisher, 2001, pp. 426–443.

[65] S. Cosentino, S. Sessa, and A. Takanishi, "Quantitative Laughter Detection, Measurement, and Classification—A Critical Survey," *IEEE Reviews in Biomedical Engineering*, vol. 9, pp. 148–162, 2016.

[66] P. Thibault, P. Gosselin, M. L. Brunel, and U. Hess, "Children's and adolescents' perception of the authenticity of smiles," *Journal of Experimental Child Psychology*, vol. 102, no. 3, pp. 360–367, 2009.

[67] P. Thibault, M. Levesque, P. Gosselin, and U. Hess, "The duchenne marker is not a universal signal of smile authenticity - but it can be learned!" *Social Psychology*, vol. 43, no. 4, pp. 215–221, 2012.

[68] X. Mai, Y. Ge, L. Tao, H. Tang, C. Liu, and Y.-J. Luo, "Eyes Are Windows to the Chinese Soul: Evidence from the Detection of Real and Fake Smiles," *PLoS ONE*, vol. 6, no. 5, p. e19903, may 2011.

[69] M. J. Bernstein, D. F. Sacco, C. M. Brown, S. G. Young, and H. M. Claypool, "A preference for genuine smiles following social exclusion," *Journal of Experimental Social Psychology*, vol. 46, no. 1, pp. 196–199, jan 2010.

[70] R. Gadassi and N. Mor, "Confusing acceptance and mere politeness: Depression and sensitivity to Duchenne smiles," *Journal of Behavior Therapy and Experimental Psychiatry*, vol. 50, pp. 8–14, mar 2016.

[71] R. Song, H. Over, and M. Carpenter, "Young children discriminate genuine from fake smiles and expect people displaying genuine smiles to be more prosocial," *Evolution and Human Behavior*, vol. 37, no. 6, pp. 490–501, nov 2016.

[72] S. Matsuda and J. Yamamoto, "Research in Autism Spectrum Disorders Intervention for increasing the comprehension of affective prosody in children with autism spectrum disorders," *Research in Autism Spectrum Disorders*, vol. 7, no. 8, pp. 938–946, 2013.

[73] J. Cockburn, M. Bartlett, J. Tanaka, J. Movellan, M. Pierce, and R. Schultz, "SmileMaze : A Tutoring System in Real-Time Facial Expression Perception

and Production in Children with Autism Spectrum Disorder," in *Intl Conference on Automatic Face and Gesture Recognition, Workshop on Facial and Bodily expressions for Control and Adaptation of Games.*, 2008.

[74] D. McDuff, R. Kaliouby, D. Demirdjian, and R. Picard, "Predicting Online Media Effectiveness Based on Smile Responses Gathered Over the Internet," in *Automatic Face and Gesture Recognition (FG), 2013 10th IEEE International Conference and Workshops on.* IEEE, 2013, pp. 1 – 7.

[75] M. B. Harms, A. Martin, and G. L. Wallace, "Facial Emotion Recognition in Autism Spectrum Disorders: A Review of Behavioral and Neuroimaging Studies," *Neuropsychology Review*, vol. 20, no. 3, pp. 290–322, 2010.

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

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

[78] D. Watson, L. A. Clark, and A. Tellegen, "Development and validation of brief measures of positive and negative affect: the PANAS scales." *Journal of personality and social psychology*, vol. 54, no. 6, pp. 1063–70, jun 1988.

[79] M. Quirin, M. Kazén, and J. Kuhl, "When nonsense sounds happy or helpless: The Implicit Positive and Negative Affect Test (IPANAT)." *Journal of Personality and Social Psychology*, vol. 97, no. 3, pp. 500–516, 2009.

[80] B. K. Payne, C. M. Cheng, O. Govorun, and B. D. Stewart, "An inkblot for attitudes: Affect misattribution as implicit measurement." *Journal of Personality and Social Psychology*, vol. 89, no. 3, pp. 277–293, 2005.

[81] P. Ekman, W. Friesen, and J. Hager, "FACS Investigator's Guide," 2002.

[82] P. Ekman and W. P. Friesen, "Measuring facial movement with the Facial Action Coding System," in *Emotion in the human face*, 2nd ed., P. Ekman, Ed. Cambridge University Press, 1982, ch. 9, pp. 178–211.

[83] J. Cohen, "Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit." *Psychological Bulletin*, vol. 70, no. 4, pp. 213–220, 1968.

[84] A. J. Viera and J. M. Garrett, "Understanding interobserver agreement:," *Family Medicine*, vol. 37, no. 5, pp. 360–363, 2005.

[85] S. L. Happy and A. Routray, "Automatic Facial Expression Recognition Using Features of Salient Facial Patches," *IEEE Transactions on Affective Computing*, vol. 6, no. 1, pp. 1–12, 2015.

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

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

[88] W. J. Yan, X. Li, S. J. Wang, G. Zhao, Y. J. Liu, Y. H. Chen, and X. Fu, "CASME II: An improved spontaneous micro-expression database and the baseline evaluation," *PLoS ONE*, vol. 9, no. 1, pp. 1–8, 2014.

[89] X. Li, T. Pfister, X. Huang, G. Zhao, and M. Pietikainen, "A Spontaneous Micro-expression Database: Inducement, collection and baseline," *2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition, FG 2013*, 2013.

[90] R. Gross, J. Shi, and J. Cohn, "Quo vadis Face Recognition?" in *Third Workshop on Empirical Evaluation Methods in Computer Vision*, no. June, 2001, pp. 119–132.

[91] G. Zhao, X. Huang, M. Taini, S. Z. Li, and M. Pietikäinen, "Facial expression recognition from near-infrared videos," *Image and Vision Computing*, vol. 29, no. 9, pp. 607–619, aug 2011.

[92] T. Pfister, X. Li, G. Zhao, and M. Pietikäinen, "Differentiating Spontaneous From Posed Facial Expressions Within a Generic Facial Expression Recognition Framework," in *IEEE International Conference on Computer Vision Workshops*, 2011, pp. 868–875.

[93] G. E. Schwartz, P. L. Fair, P. Salt, M. R. Mandel, and G. L. Klerman, "Facial expression and imagery in depression: an electromyographic study." *Psychosomatic medicine*, vol. 38, no. 5, pp. 337–47, 1976.

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

[95] D. Matsumoto, P. Ekman, and A. Fridlund, "Analyzing Nonverbal Behavior," in *Practical guide to using video in the Behavioral Sciences*, P. W. Dowrick, Ed. John Wiley & Sons, Inc., 1991, ch. 10, pp. 153–165.

[96] M. Hamedi, S.-H. Salleh, M. Astaraki, and A. M. Noor, "EMG-based facial gesture recognition through versatile elliptic basis function neural network." *Biomedical engineering online*, vol. 12, no. 1, p. 73, 2013.

[97] M. Perusquía-Hernández, M. Hirokawa, and K. Suzuki, "A wearable device for fast and subtle spontaneous smile recognition," *IEEE Transactions on Affective Computing*, vol. 8, no. 4, pp. 522–533, 2017.

[98] S. Koelstra, C. Muhl, M. Soleymani, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras, "DEAP: A Database for Emotion Analysis ;Using Physiological Signals," *IEEE Trans. Affect. Comput.*, vol. 3, no. 1, pp. 18–31, 2012.

[99] S. M. Mavadati, M. H. Mahoor, K. Bartlett, P. Trinh, and J. F. Cohn, "DISFA: A spontaneous facial action intensity database," *IEEE Transactions on Affective Computing*, vol. 4, no. 2, pp. 151–160, 2013.

[100] S. Wang, Z. Liu, Z. Wang, G. Wu, P. Shen, S. He, and X. Wang, "Analyses of a multimodal spontaneous facial expression database," *IEEE Transactions on Affective Computing*, vol. 4, no. 1, pp. 34–46, 2013.

[101] M. K. Abadi, R. Subramanian, S. M. Kia, P. Avesani, I. Patras, and N. Sebe, "DECAF: MEG-Based Multimodal Database for Decoding Affective Physiological Responses," *IEEE Transactions on Affective Computing*, vol. 6, no. 3, pp. 209–222, 2015.

[102] S. Porter and L. Ten Brinke, "Reading Between the Lies," *Psychological Science*, vol. 19, no. 5, pp. 508–514, 2008.

[103] W.-j. Yan, W. Qi, L. Jing, Y.-h. Chen, and X. Fu, "How Fast are the Leaked Facial Expressions : The Duration of Micro-Expressions," *Journal of Nonverbal Behavior*, vol. 37, no. 4, pp. 217–230, 2013.

[104] X.-b. Shen, Q. Wu, and X.-l. Fu, "Effects of the duration of expressions on the recognition of microexpressions," *Journal of Zhejiang University SCIENCE B (Biomedicine & Biotechnology)*, vol. 13, no. 3, pp. 221–230, 2012.

[105] M. Shreve, S. Godavarthy, D. Goldgof, and S. Sarkar, "Macro- and micro-expression spotting in long videos using spatio-temporal strain," *2011 IEEE International Conference on Automatic Face and Gesture Recognition and Workshops, FG 2011*, pp. 51–56, 2011.

[106] X. Ben, P. Zhang, R. Yan, M. Yang, and G. Ge, "Gait recognition and micro-expression recognition based on maximum margin projection with tensor representation," *Neural Computing and Applications*, 2015.

[107] F. Xu, J. Zhang, and J. Wang, "Microexpression Identification and Categorization using a Facial Dynamics Map," *IEEE Transactions on Affective Computing*, vol. PP, no. 99, pp. 1–1, 2016.

[108] X. Huang, G. Zhao, X. Hong, W. Zheng, and M. Pietikäinen, "Spontaneous facial micro-expression analysis using Spatiotemporal Completed Local Quantized Patterns," *Neurocomputing*, vol. 175, pp. 564–578, 2015.

[109] Z. Xia, X. Feng, J. Peng, X. Peng, and G. Zhao, "Spontaneous micro-expression spotting via geometric deformation modeling," *Computer Vision and Image Understanding*, vol. 147, pp. 87–94, 2015.

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

[111] Y. Takano and K. Suzuki, "Affective communication aid using wearable devices based on biosignals," in *Proceedings of the 2014 conference on Interaction design and children - IDC '14*. New York, New York, USA: ACM Press, 2014, pp. 213–216.

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

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

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

[115] J. L. Fleiss, "Measuring nominal scale agreement among many raters," *Psychological Bulletin*, vol. 76, no. 5, pp. 378–382, 1971.

[116] C. Amma, T. Krings, J. Böer, and T. Schultz, "Advancing Muscle-Computer Interfaces with High-Density Electromyography," *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15*, pp. 929–938, 2015.

[117] W.-J. Yan, Q. Wu, J. Liang, Y.-H. Chen, and X. Fu, "How Fast are the Leaked Facial Expressions: The Duration of Micro-Expressions," *Journal of Nonverbal Behavior*, vol. 37, no. 4, pp. 217–230, 2013.

[118] IPanda, "Panda wants a hug from nanny, but nanny is working," 2017.

[119] A. Brown, "Trololo cat," 2010.

[120] P. Lang, M. Bradley, and B. Cuthbert, "International Affective Picture System (IAPS)," University of Florida, Gainesville, FL., Tech. Rep., 2008.

[121] S. Shimoda, N. Ōkubo, M. Kobayashi, S. Satō, and H. Kitamura, "An attempt to construct a japanese version of the implicit positive and negative affect test (ipanat)," *Shinri-gaku kenkyū*, vol. 85, no. 3, pp. 294–303, 2014.

[122] D. Girardi, F. Lanubile, and N. Novielli, "Emotion Detection Using Noninvasive Low Cost Sensors," in *Affective Computing and Intelligent Interaction*, 2017, pp. 125–130.

[123] S. Taylor, N. Jaques, Weixuan Chen, S. Fedor, A. Sano, and R. Picard, "Automatic identification of artifacts in electrodermal activity data," in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, aug 2015, pp. 1934–1937.

[124] H. Beyer and K. Holtzblatt, "Contextual design," *interactions*, vol. 6, no. 1, pp. 32–42, jan 1999.

[125] C. A. Meissner and J. C. Brigham, "Thirty Years of Investigating the Own-Race Bias in Memory for Faces: A Meta-Analytic Review," *Psychology, Public Policy, and Law*, vol. 7, no. 1, pp. 3–35, 2001.

[126] K. L. Hourihan, A. S. Benjamin, and X. Liu, "A cross-race effect in metamemory: Predictions of face recognition are more accurate for members of our own race." *Journal of applied research in memory and cognition*, vol. 1, no. 3, pp. 158–162, sep 2012.

[127] J. Kruger and D. Dunning, "Unskilled and Unaware of It: How Difficulties in Recognizing One's Own Incompetence Lead to Inflated Self-Assessments," *Journal of Personality and Social Psychology*, vol. 77, no. 6, pp. 1121–1134, 1999.

[128] M. T. Orne, "Demand Characteristics and the Concept of Quasi-Controls1," in *Artifacts in Behavioral Research*. Oxford University Press, may 2009, pp. 110–137.

# Acknowledgments

This thesis is the last step to complete the Doctorate in Empowerment Informatics at the University of Tsukuba. It is the result of several years of hard work, not only mine but also from a group of people who collaborated with me to make this achievement possible. During these years, I worked, not only in the topics covered in this dissertation, but also on a side project that was an important part of my work during the Ph.D.: Levitas. Hereby I would also like to thank the collaborators involved in that project.

First of all, I would like to thank my wonderful supervisors for their inspiring ideas, support, patience, and enthusiasm. To Professor Kenji Suzuki, for all the time he invested in guiding me throughout these difficult years. Without your support and constant encouragement this would not have been possible. Thank you very much for trusting that my ideas were worth implementing, even if they were once only paper. To Professor Saho Ayabe, for all the regular meetings, her enthusiasm, and thoughtful advice to make of this work sound in the field of Psychology. To Professor Hideaki Kuzuoka, for the stimulating discussions we held and that enriched this work. To Professor Mai Otsuki, for all her guidance during the Levitas project, and the always interesting questions she posed to challenge me. To Professor Kazuya Inoue, for all his efforts to discuss psychology topics in English with me, and for his support preparing the experiment on human judgment's stimuli presentation. To Professor Fumiaki Murakami for his advice regarding art and design. Finally, thanks to professor Hiroo Iwata, for his kind support during the Levitas project.

Also, I would like to thank the Empowerment Faculty who supported me during these years. Professor Masakazu Hirokawa, thank you for co-authoring some of the papers presented in this thesis, and for the discussions that made them possible. Your guidance and moral support throughout the submissions were invaluable to me and contributed greatly on the success of this journey. Professor Jun Izawa, I could not have enjoyed more your lectures on Neuromotor Control. Thank you for sharing with us, students, your enthusiasm and passion for your research. Your lectures and our conversations during the Lab Rotation inspired me greatly at the beginning of

this work. My dear Professor Aki Yamada, thank you for being an example of both strength and kindness, intelligence and dedication for your students. Your support and advice made it possible for me to expand my horizons not only in Japan, but also to China and the US.

I would like to thank my student collaborators, who made it possible and enjoyable the long hours of development and data collection in the different projects. Hizashi Kaeri, thank you for your support conducting experiment 2 and 3 in Japanese and Chinese. Also thanks for the long funny conversations we had in Japanese. It was you the first person that really encouraged me to use this language. For the Levitas project, Tiago Martins, thanks for your encouragement, creative, and technical input on the project, even at long distance. Without you it would not have been possible. Takahisa Enomoto thank you for your enthusiasm in the development of the Motion Base server, and your patience to conduct the necessary experiments. Akiko Hirai, I enjoyed very much working with you. Thank you for your speed improving the Motion Base server setup, and for your sharp eye to troubleshoot problems.

My sincere gratitude to all the members of the Artificial Intelligence Laboratory who supported me in the everyday hassles of conducting research and living in Japan. Specially I would like to thank Airi Tsuji for the long hours she spent video-coding the data from experiment 1; Baptiste Bourreau for his enthusiasm developing a custom version of Bioroid to be used during experiment 2; and to Dushyantha Jayatilake and Keisuke Kawahara for their advice regarding hardware design and implementation for the Levitas project.

Also, I thank the EMP fellow students, especially Masa Jazbec, for their friendship and support during these years, and to the EMP office staff for all their administrative support. Moreover, I would like to thank the hundreds of participants that made this research possible, and to Monserrat Corona who kindly proofread this work.

Finally, thanks to my family, whose support has been invaluable. Especially to my sisters, Carolina and Karina, who kindly modeled and took some of the pictures for this thesis.

Gracias a ti, la luz más brillante de mi universo. Tan ardientemente enceguecedora, que fue capaz de traerme la felicidad más grande, pero también de proyectar la sombra más profunda en la que me he sumergido jamás. Gracias a ti, y a pesar de ti, he terminado.

# About

Monica Perusquía-Hernández received the B.Sc. degree in Electronic Systems Engineering from the Instituto Tecnologico y de Estudios Superiores de Monterrey, Mexico, in 2009; and the M.Sc. degree in Human-Technology Interaction and the Professional Doctorate in Engineering in User-System Interaction from the Eindhoven University of Technology, the Netherlands, in 2012 and 2014, respectively. Currently she is Ph.D. student at the Empowerment Informatics Program, University of Tsukuba, Japan. Her research interests include Affective computing, Biosignal processing, Augmented Human Technology, and Artificial Intelligence.

### Publications

### Journal Paper

1. Perusquía-Hernández, M., Hirokawa, M., Suzuki, K., A wearable device for fast and subtle spontaneous smile recognition. IEEE Transactions on Affective Computing Vol. 8, no. 4, pp. 522-533. 2017.

2. Severens, M., Perusquía-Hernández, M., Nienhuis, B., Farquhar, J., Duysens, J., Using Actual and Imagined Walking Related Desynchronisation Features in a BCI, IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 23, issue 5, pp.877-886. 2014.

### Conference Paper

1. Perusquía-Hernández, M., Hirokawa, M., Suzuki, K., "Spontaneous and posed smile recognition based on spatial and temporal patterns of facial EMG", Proceedings of the 7th Affective Computing and Intelligent Interaction Conference, pp. 537-541, 4 pages. 2017.

2. Perusquía-Hernández, M., Enomoto, T., Martins, T., Otsuki, M., Iwata, H., Suzuki, K., "Embodied interface for levitation and navigation in a 3D large

space", Proceedings of the 8th Augmented Human International Conference. Article 4, 8 pages. 2017.

3. Perusquía-Hernández, M., Martins, T., Enomoto, T., Otsuki, M., Iwata, H., Suzuki, K., "Multimodal Embodied Interface for Levitation and Navigation in 3D Space", Proceedings of the 2016 Symposium on Spatial User Interaction, pp. 215, 1 page. 2016.

4. Perusquía-Hernández, M., Chen, W., Feijs, L., "Garment Design for an Ambulatory Pregnancy Monitoring System. Ambient Assisted Living and Daily Activities", Proceedings of the 6th International Work-Conference, IWAAL 2014, pp. 219-227, 8 pages. 2014.

5. Perusquía-Hernández, M., Kriening, H., Palumbo, C., Wajda, B., "User-centered design of a lamp customization tool", Proceedings of the 5th Augmented Human International Conference, Article 36, 2 pages. 2014.

**Book chapter**

1. Perusquía-Hernández, M., Chen, W., Feijs, L., "Textile-integrated electronics for ambulatory pregnancy monitoring". Book chapter in Advances in smart medical textiles - Woodhead Publishing. Pages 239–268. 2016.

**Scientific presentations**

1. Perusquia-Hernandez, M., Severens, M., Farquhar, J., Cuijpers, R.H., "A Brain-Computer Interface for Walking". Poster presentation. BBCI Workshop 2012 on Advances in Neurotechnology. 2012.

2. Perusquia-Hernandez, M., Suzuki, K., "A wearable device for fast and subtle spontaneous smile recognition" at the Nichibokubashi Symposium at the Mexican Consulate.