

FAMOS: a framework for investigating the use of face features to identify spontaneous emotions

DA COSTA ABREU, Marjory <<http://orcid.org/https://orcid.org/0000-0001-7461-7570>> and BEZERRA, G.S.

Available from Sheffield Hallam University Research Archive (SHURA) at:
<http://shura.shu.ac.uk/25391/>

This document is the author deposited version. You are advised to consult the publisher's version if you wish to cite from it.

Published version

DA COSTA ABREU, Marjory and BEZERRA, G.S. (2019). FAMOS: a framework for investigating the use of face features to identify spontaneous emotions. *Pattern Analysis and Applications*, 22 (2), 683-701.

Copyright and re-use policy

See <http://shura.shu.ac.uk/information.html>

FAMOS: A framework for investigating the use of face features to identify spontaneous emotions

Abstract—Emotion-based analysis has raised a lot of interest, particularly in areas such as forensics, medicine, music, psychology, and human-machine interface. Following this trend, the use of facial analysis (either automatic or human-based) is the most common subject to be investigated once this type of data can easily be collected and is well accepted in the literature as a metric for inference of emotional states. Despite this popularity, due to several constraints found in real world scenarios (e.g. lightning, complex backgrounds, facial hair and so on), automatically obtaining affective information from face accurately is a very challenging accomplishment. This work presents a framework which aims to analyse emotional experiences through spontaneous facial expressions. The method consists of a new 4-dimensional model, called FAMOS, to describe emotional experiences in terms of appraisal, facial expressions, mood, and subjective experiences using a semi-automatic facial expression analyser as ground-truth for describing the facial actions. In addition, we present an experiment using a new protocol proposed to obtain spontaneous emotional reactions. The results have suggested that the initial emotional state described by the participants of the experiment was different from that described after the exposure to the eliciting stimulus, thus showing that the used stimuli were capable of inducing the expected emotional states in most individuals. Moreover, our results pointed out that spontaneous facial reactions to emotions are very different from those in prototypic expressions, especially in terms of expressiveness.

Index Terms—Facial expression analysis, Face biometrics, Emotion analysis, Action units, Emotion models.

1 INTRODUCTION

Emotion plays an important role in daily life due to its effect on people's behaviour. Several studies have investigated the physiological changes related to emotion to find a correspondence between emotional states and biometrics. To do so, automatic systems are designed for extracting individual physiological measures (e.g. electrodermal activity, heart rate, brain's electrical activity) obtained in response to outer stimuli [1], [2], [3], [4], [5], [6], [7], [8], [9]. These studies have shown that the collected measures are similar in people experiencing the same emotion, thereby allowing the inference of emotional states based on the interpretation of physiological reactions. For example, the polygraph is a tool widely used nowadays to identify physiological reactions to emotion and, as a result, it works as an efficient lie detector, although its accuracy is penalised by the cost factor [10].

Aiming of finding a cheaper method and use it in natural human interactions, facial-based systems have been developed [11], [12], [13], [14], [15], [16], [17], [18], [19], [20], [21], [22]. A general system consists of face detection, extraction of facial features, expression recognition, and emotion inference [23]. Developing such automated facial expression analysis system is a complex task because there are many factors which affect the processing of facial information. Face shape, facial hair, eyeglasses, age, gender, occlusion, differences in expressiveness, and degree of facial plasticity are examples of common constraints which must be considered by

robust algorithms to obtain consistent results. To do so, the automated system requires sufficient information to handle this wide spectrum of ethnic and physiological factors. In addition, the system have to operate at real time whereas facial responses might be fast and sudden. There are many methods to perform each step of the facial expressions analysis depending on constraints in face acquisition (e.g. image resolution, illumination, and occlusion) and the desired efficiency in the recognition task (e.g. real-time responses and error rate).

The interest of studying facial expressions in order to recognise emotion is not new and there is a wide spectrum of applications for this specific aim, such as prediction of soft-biometrics or health-care, human machine interaction, diagnoses, lie detection, and so on [24]. The knowledge of the individual's emotional state provides information which could be extremely valuable in interpreting particular scenarios or evaluating human activities. For obtaining such knowledge, it is important to understand why and how we react to emotions through facial expressions.

Darwin (1872) was the first who claimed that emotions and their expressions were biologically innate and evolutionarily adaptive. Ekman and Friesen (1971) conducted a study to investigate the universality of emotion facial expressions, whose findings led to the conclusion that people have the innate ability to perform and interpret six facial expressions (*happiness, anger, disgust, fear, surprise, and sadness*), which are called universal, although their intensity and initiation depend on cultural factors [25]. A recent study has suggested four basic emotions instead of six [26]. This conclusion was reached by the analysis of each facial muscle activated

in signalling emotions which showed clearly differences between facial expressions of *happiness* and *sadness*, while similarities between other emotions were found (*fear* and *surprise*; *anger* and *disgust*). Despite the different perspectives on the universality of some emotion facial expressions, there is a general representation of them, which is called prototypic. Most studies encompassing facial expression analysis take prototypic expressions as basis for training, causing a bias that does not present emotion facial expression analysis in real-world scenarios. For this reason, contemporaneous studies have been considering spontaneous facial expressions [27], [28], [29], [30], [31], [32], [24], [33], but there is still unresolved issues in this field.

Although the prototypic expressions of basic emotions are universal [34], which makes them easy to be identified, natural expressions (expressions elicited spontaneously in daily life) differ mainly in expressiveness, which is not universal [34] and therefore they can lead us to deduce incorrect emotional states.

Some studies have shown the impact of natural and prototypical expressions on the accuracy of classifiers [35], [36], [21]. The results showed a higher accuracy rate when using a database with prototypic facial expressions than using spontaneous ones. For example, Valstar and Pantic [21] obtained a recognition rate of 72% using spontaneous facial expressions.

In this paper we investigate the use of spontaneous facial expressions in emotion analysis to explore the possibility of developing a system able to infer emotional states, and to determine the extent to which this type of personal characteristic inference is likely to be possible in practice. We have investigated spontaneous facial expressions over prototypic ones so that we could find a system close to human interaction which included: sonorous and visual stimuli, facial expression analysis, and emotion inference based on newly collected data.

There are many reasons which encourage the development of an emotion analysis method considering spontaneous facial expressions:

- Emotion analysis can be applied in different contexts providing essential information which could be helpful for areas such as medicine [37], [38], [39], psychology [40], forensics [41], human-machine interface [42], music [43], graphic animations [44], and so on.
- As previously mentioned, most studies explore prototypic facial expressions over spontaneous ones and therefore the resultant methods are not suitable for real-world scenarios. Thus, the analysis of spontaneous facial expressions could lead to the development of more robust approaches in facial expression analysis.
- The use of facial features in emotion analysis is a subject addressed by theorists who discuss the universality of emotions. The existence of similar facial reactions to emotions across different cultures is widely explored in the literature in attempting

to prove or disprove this hypothesis, which once disproved could lead to different perspectives on emotion inference from face.

This paper is organised as follows: Section 2 describes the emotion models from the literature which were considered during the elaboration of the FAMOS; Section 3 presents a semi-automatic facial expression analyser employed to identify facial expressions; The stimuli for eliciting spontaneous facial expressions and the database acquisition process are presented in Section 4; Section 5 discusses the obtained results in each experimental scenario; and finally, Section 6 shows our conclusions and perspectives for future work.

2 EMOTION MODEL

Emotion models are intended for recognizing emotions considering their features. A well defined emotion model is essential to cover the main parameters related to the affective experience. Among different emotion models, two-dimensional's, where emotions are arranged as a valence-arousal vector, are the most common [45], [46], [47], [8]. Despite this, there is an increasingly growth of interest in exploring new dimensions, since two dimensions are not always enough to describe affective experiences because they are not capable of representing the rich semantic space of emotion [48], [49]. Besides, there are studies which have pointed out that the arousal dimension should not be considered as an atomic measure whereas it is composed of two sub-dimensions, arousal–calmness and tension–relaxation, related to opposite causes [50]. This perspective has encouraged the use of multidimensional models with more than the two traditional dimensions in recent studies [51], [52], [53].

In attempting to support the multidimensional approach, a study demonstrated that emotion models should consider a set of six emotion components, namely *appraisal of events*, *psychophysiological changes*, *motor expressions*, *action tendencies*, *subjective experiences*, and *emotion regulation*, due to their high correlation with emotion experiences related to 24 prototypical emotion terms [48]. The four dimensions highlighted in this study were those which presented the greatest variance in the analysed sample: *evaluation-pleasantness*, *potency-control*, *activation-arousal*, and *unpredictability*. The dimensions were showed to be significantly correlated to the six emotion components, which in turn have been shown to describe the emotional experience robustly [54], [55]. These results also imply that simple two-dimensional models, which are common in the literature, miss major sources of variation in the emotion domain.

Taking these results into account, these six emotion components were considered as a reference for the elaboration of a new multidimensional model called FAMOS (Facial expressions, Appraisal, MOod, and Subjective experiences). The emotion words explored in this model were those related to three basic emotions: *happiness*, *sadness* and *fear*. They were selected due to the possibility

of provoking such emotions through aesthetic art forms (music and images) strongly enough to produce visible physical changes and without manipulating people [56]. For this reason, *anger/disgust* were not explored in this study.

Figure 1 illustrates each of these dimensions showing how they were obtained. The input consists of both subjective and physiological factors (facial expressions). Each subjective dimension is described as a numerical attribute scaled from 1 to 5, obtained from questionnaires, while the facial expressions are described by action units (AUs), obtained by a Facial Expression Analyser (FEA). As a result, it is possible to describe the emotion experienced in terms of these dimensions grouping the output data into basic emotions.

The FAMOS consists of the following dimensions:

- **Mood (action tendencies component):** It represents the individual's emotional state before their exposure to the eliciting stimulus. Subjects describe their initial emotional state in a 1-5 Likert scale, being 1 related to low valence emotions (*sadness*), 3 related to distress feelings (*fear*), and 5 related to high valence emotions (*happiness*). It is also possible to choose a middle ground between either *sadness* and *fear* or *fear* and *happiness*, depending on the subject's perspective on their mood. The mood can influence the arousing emotion process [57] and therefore it was considered in the emotional experience analysis in order to find its correlation with the reported appraisal.
- **Appraisal (appraisal of events component):** It represents the appraisal associated with the eliciting stimulus. After being exposed to such stimulus, the individual is asked to answer a second questionnaire, where they inform their appraisal on a 1-5 Likert scale, being 1 related to low valence emotions (*sadness*), 3 related to distress feelings (*fear*), and 5 related to high valence emotions (*happiness*). It is also possible to choose a middle ground between either *sadness* and *fear* or *fear* and *happiness*, depending on the subject's perspective on their emotional state. These ratings are used to aid the stimulus efficiency assessment by checking whether the appraisal is consistent with the expected emotion. Furthermore, the correlation between subjective reports and facial responses is evaluated in attempting to find empirical evidence about the agreement between them.
- **Subjective experiences (subjective experiences component):** Episodic memory can produce several different types of emotion in a short period of time [57], which makes the knowledge of personal information (e.g. previous contact with the eliciting stimulus) from an individual highly desirable [56]. Concerning the auditory stimuli, individuals who have musical experience, for example, could perceive hidden details that can help the cognitive process of understanding musical cues, which is necessary to arouse emotional experiences through music.

We also considered people who easily get emotional, once the emotional trigger time is short, and the eclectic ones, who are more receptive to musical cues. Taking these facts into account, this personal information was collected through the application of questionnaires after and before the exposure to the eliciting stimuli.

- **Facial expressions (motor component):** In response to an emotion eliciting situation, a person can demonstrate an emotion in many ways and one of the most significant is through facial expressions. Mehrabian [58] indicated that the verbal stimuli (i.e., spoken words) of a message contributes only with 7% of the effect of the message as a whole, the vocal stimuli (e.g., voice intonation) contributes with 38%, while facial expression of the speaker contributes with 55% of the spoken message effect. Thus, it is possible to infer the emotion experienced by someone by means of the analysis of their facial expressions. In this work, the facial expressions were obtained through facial pictures taken from filming people during their exposure to an eliciting stimulus. Then, a FEA was employed to extract AUs from each picture, as will be presented in Section 3, which subsequently allowed to check the existence of AUs patterns associated with the basic emotions.

We have adopted Likert scales for representing Mood and Appraisal since they are more intuitive and easier to rate by individuals than usual 2-D valence-arousal vectors. In addition, the arousal component is already tied to the emotion terms (e.g. Happiness: high valence, average arousal [59]).

The subjective dimensions explored in this work were associated with the processing of musical cues in attempting to check links between musical load and emotional experiences. The ratings from individuals were compared to their facial reactions, which were extracted by a FEA. The results obtained by the FEA were compared to human assessment as well, in order to evaluate the accuracy and precision of the method.

The emotion components *emotion regulation* and *psychophysiological changes* were not explored in this study since the emotions were assumed to be spontaneous (without attenuation, amplification, concealment, and substitution), and only motor expressions were considered in attempting to obtain a low-cost method (the physiological data acquisition is quite expensive).

Few studies have considered contextual information to improve the facial expression analysis task [32], therefore we have analysed the effect of subjective information on the emotional experience for this particular goal.

3 FACIAL EXPRESSION ANALYSIS

Facial expressions can be defined as changes in the face provoked by internal emotional states in response to external stimuli or to communicate the individual's intentions [60]. Processing information from face has

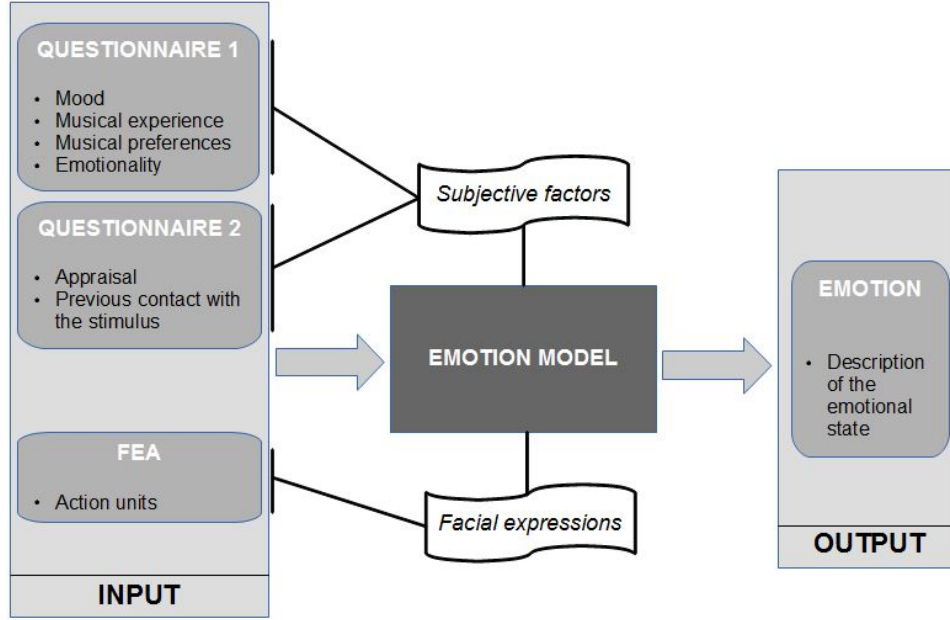


Fig. 1: FAMOS emotion model design.

helped humans to survive since the time when there was no language. For instance, perceiving someone's intentions was fundamental to escape from predators or to recognise potential enemies and reproductive partners [61]. This ability is more prominent in women, who are most able to process facial information according to studies in this field [62], [63], which suggest that the reason for these results is the women's need to look after of infants by decoding and detecting distress on their faces or in order to protect them against threatening signals from other individuals [64].

In order to describe facial expressions, Ekman and Friesen [65] developed a facial code called Facial Action Coding System (FACS). The system describes the physical expression of emotions through AUs, which are the smallest visually discernible facial movements related to each facial muscle or group of facial muscles. Table 1 shows the upper and lower face AUs used in our study.

In order to find an unified set of AUs to describe spontaneous emotions, we have employed a method based on a model-driven technique, which depends on prior information about the faces (a neutral face) and landmark detection. The presented method infers the AUs based on the differences obtained between neutral and expressive faces. We have chosen a local model-based method since it has the advantage of being suitable to both single images and image sequences and it is not affected by age wrinkles since a template image is taken into account during the inference of facial expressions. Also, it does not require extensive prior knowledge about the object of interest, which is required by image-based approaches. The proposed FEA is illustrated in Figure 2.

Each step of the facial feature extraction can be described in details as follows:

- 1) **Face detection:** It is responsible for extracting the

TABLE 1: Upper and lower face AUs [65].

| AU | Muscular Basis |
|---------------------------|--|
| 0 - Neutral Face | - |
| 1 - Inner Brow Raiser | <i>Frontalis, Pars Medialis</i> |
| 2 - Outer Brow Raiser | <i>Frontalis, Pars Lateralis</i> |
| 4 - Brow Lowerer | <i>Depre. Glabe.; Depre. Super.; Corrug.</i> |
| 5 - Upper Lid Raiser | <i>Levator Palpebrae Superioris</i> |
| 6 - Cheek Raiser | <i>Orbicularis Oculi, Pars Orbitalis</i> |
| 7 - Lid Tightener | <i>Orbicularis Oculi, Pars Palpebralis</i> |
| 9 - Nose Wrinkler | <i>Levator Labii Superioris, Alaeque Nasi</i> |
| 10 - Upper Lip Raiser | <i>Levator Labii Superioris, Caput Infra.</i> |
| 12 - Lip Corner Puller | <i>Zygomatic Major</i> |
| 15 - Lip Corner Depressor | <i>Triangularis</i> |
| 16 - Lower Lip Depressor | <i>Depressor Labii</i> |
| 17 - Chin Raiser | <i>Mentalis</i> |
| 20 - Lip Stretcher | <i>Risorius</i> |
| 23 - Lip Tightener | <i>Orbicularis Oris</i> |
| 24 - Lip Pressor | <i>Orbicularis Oris</i> |
| 25 - Lips Part | <i>Depressor Labii, Mentalis, Orbicularis Oris</i> |
| 26 - Jaw Drop | <i>Massetter; Temp. and Inter. Pterygoid Rel.</i> |
| 27 - Mouth Stretch | <i>Pterygoids; Digastric</i> |

face from the image. For this purpose, we have used the Viola-Jones method [66], which can be described by three key concepts: integral image generation, selection of Haar-like features by Adaboost algorithm and generation of the cascade classifiers. The first step of the implemented algorithm is the contrast adjustment by using Histogram Equalisation [67]. Then, the classifier is loaded through a file containing a decision tree trained by using several positive and negative images of face. The OpenCV

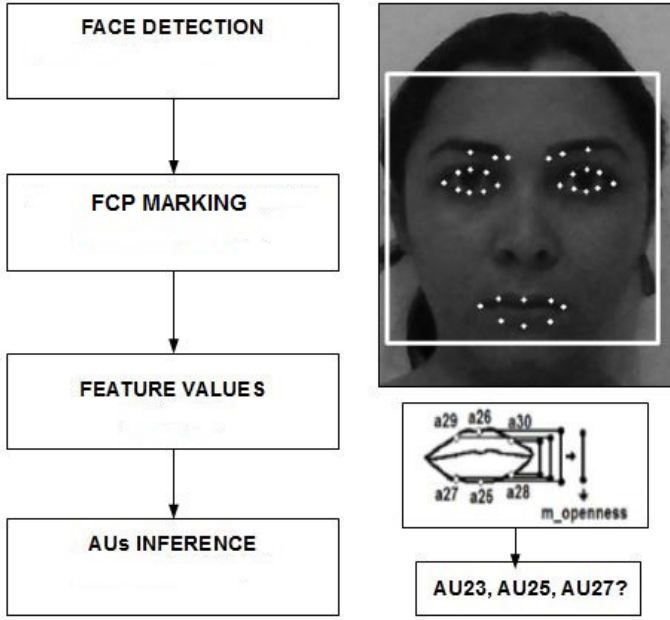


Fig. 2: Facial expression analyser.

platform [68] offers these files freely, therefore it was not necessary to train the classifier. Once the face is detected, the image is cropped removing unessential information. Figure 3 shows examples of the detection performed by this algorithm. Even for images with poor lighting, the face was detected.

- 2) **FCP marking:** Facial points are marked manually based on the geometrical model proposed by Kobayashi and Hara [11] for extracting facial features. The proposed geometrical model describes the face through 30 points called facial characteristic points (FCP). The points were chosen based on the key locations of each facial deformation which constitutes expressive facial expressions. As a result, it is possible to identify expressive changes in face by analysing the FCP obtained from each facial image. The FCP were manually marked to ensure the required accuracy in identifying subtle facial expressions, since such accuracy could be harmed by using automatic extraction methods. Finally, the facial points are normalised by applying three transformations to the coordinates of the FCP: translation, rotation and scaling.
- 3) **Calculation of feature values:** Using the normalised FCP (obtained in Step 2), the feature values were calculated by identifying geometric features on the face [13], as shown in Table 2. For example, *ieb_height* describes the inner eyebrow height, which is a feature value which considers the normalized FCP 19, 20, 21, and 22. In the original method, only the one-side upper facial points were

considered. Moreover, the calculation of geometric features was not efficient in specific situations, for example, in case of changes in eyebrows related to AUs 1 and 2, the feature value *ieb_height* contained both outer and inner brow height and therefore these AUs were always found together. Another problem was found in the feature *m_mos* which was not enough to describe some lower face AUs (e.g. AU15). In order to handle these problems, three new features have been proposed to describe these changes accurately: *ieb_height*, *oeb_height* and *lc_height*. The reformulated model is shown in Table 2.

- 4) **Action units inference:** After the calculation of feature values from neutral and expressive pictures, the obtained difference was matched to AUs by using a rule-based system. The rules were created by evaluating changes in mouth, eyes, and eyebrows presented in expressive expressions. Table 3 shows the changes in a neutral face and the corresponding AUs. Thresholds (values expressed in pixels) were obtained by training and used to consider changes in facial expressions which are not significant enough to generate a new expression. For instance, AU 1 is found if there is an increase of the inner eyebrow height, and the FCPs 19 and 20 compared to the values obtained from the neutral face.

The classifier was trained using 40 positive samples containing spontaneous facial expressions (obtained during an experiment, as will be presented in Section 4) and 40 negative samples containing neutral expressions. The changes in feature values found due to FCP normalisation were measured in order to find thresholds for each equation presented in Table 3. Afterwards, new images were submitted to the classifier and the results were compared to human assessments.

4 NEW EMOTION DATABASE ACQUISITION: TECHNICAL SPECIFICATIONS

4.1 Subjects

Since it is rare to find a freely available database with non-prototypic facial expressions, with certain exceptions [69], [70], [71], as part of our contribution, a new database with facial expressions of elicited emotions was built. The users who took part in the data collection were students from *Universidade Federal do Rio Grande do Norte* (UFRN) and employees from *SIG Software* company, considering only individuals without facial deformities, 52 females and 49 males aged 18-60. The average time spent per person during the experiment was seven minutes. Each participant was exposed to only one emotional stimulus since after the procedure the majority of people became aware of the nature of the research.

¹Source: <http://biometrics.idealtest.org/>. Access date: 04 April 2015.

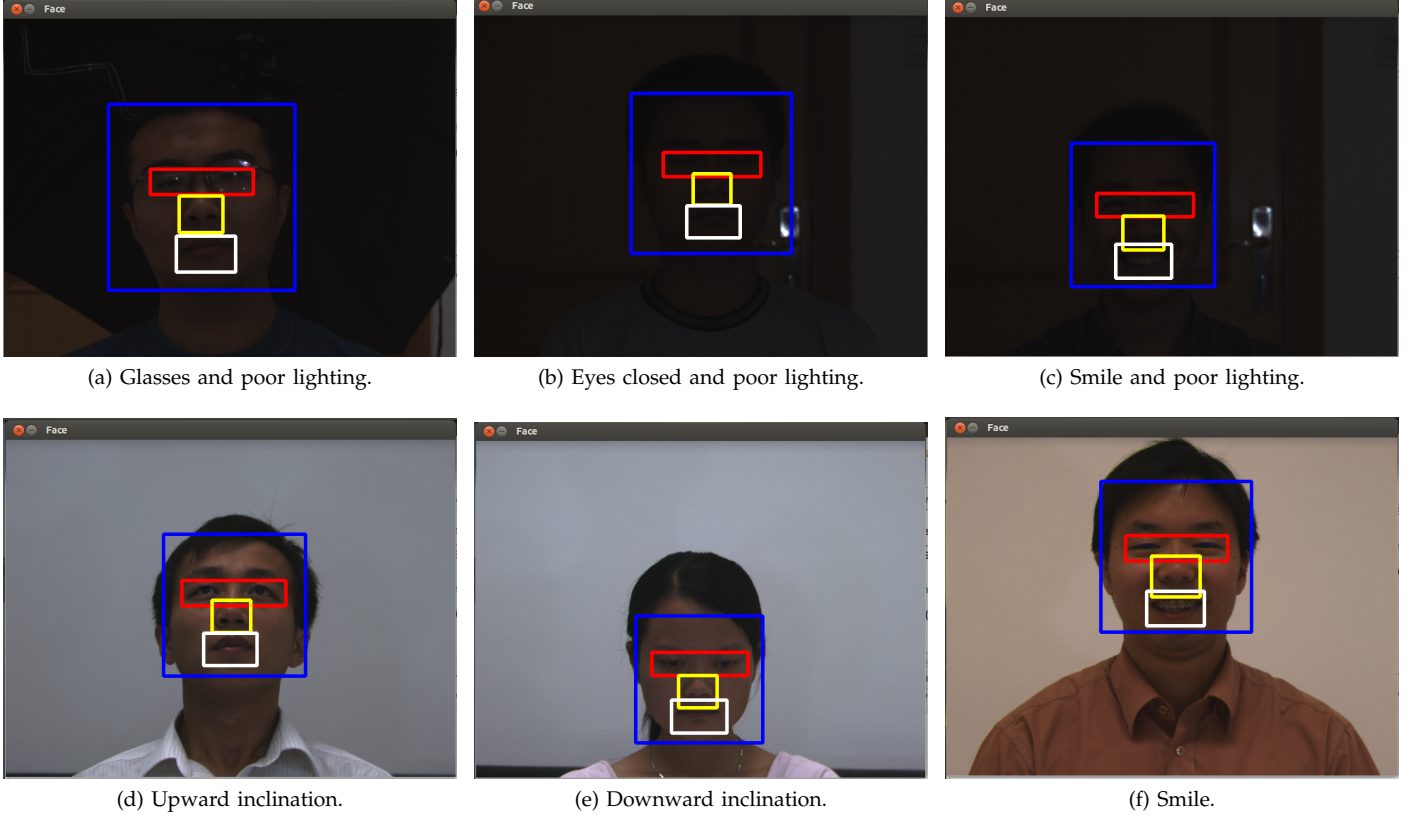


Fig. 3: Object detection using Viola-Jones method. The blue square represents the face classifier and the others represent facial features classifiers (red for eyes, yellow for nose and white for mouth)¹.

4.2 Apparatus and stimulus materials

The images were collected using a webcam (1280 x 720). Instead of analysing videos, since we have proposed a frame-based method for extracting facial features, we have captured frames from the captured data, 5 frames per second. The seven more expressive frames from each subject were used in the data analysis stage in attempting to avoid attenuation of the results provoked by excessive neutral frames.

In music, several studies have explored the effects of musical parameters on the individual's mood [1], [72], [6], [73], [74]. A survey of Music Information Retrieval (MIR) is presented by Hu (2010) who states the existence of mood effect in music [75] and shows the main musical parameters with some correspondence with listeners' judgements and divergences in the subjective factor. Taking these facts into account, we have proposed sonorous stimuli for eliciting emotional states. The musical tracks were selected based on previous experiments from the literature [1], [76], [77] that showed related high emotion levels by measuring physiological changes (e.g. heart rate, blood pressure, skin conductance level, finger temperature and respiration measures) in individuals who were listening to these tracks while they were being monitored. The selected tracks are illustrated in Table 4.

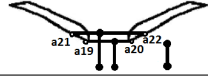
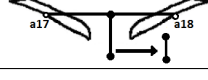
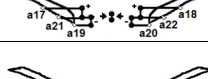
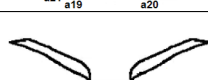


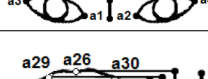
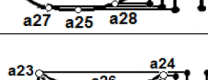


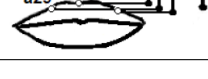
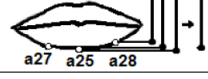
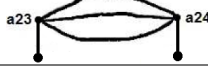
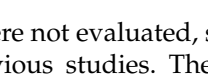
At first, the experiment was made only using the

auditory stimuli presented in Table 4, but the results showed that these stimuli were not sufficient to provoke visible changes in facial expressions and the absence of a fixed spot to focus the individual's vision generated distraction which harmed the emotion eliciting process. Taking these preliminary results into account, it was decided to employ visual stimuli as well. Verbal messages were not used, thus the eliciting stimulus could be presented to anyone without relying on linguistic factors.

Facial pictures can elicit emotion based on intrinsic human features: mimicry, empathy and cognitive load [78]. On account of this, we have selected people's images in particular emotional scenarios, depending on the emotion expected to be provoked. The pictures were chosen considering cultural categories which are related to the induction of particular emotions. For example, pictures of people smiling were chosen for eliciting *happiness*, pictures of people crying, alone, ill, victims of violence, and living in poverty were chosen for eliciting *sadness*, and pictures taken from horror movies were chosen for eliciting *fear*. Figure 4 shows samples of the used images labelled by emotion.

The images selected were submitted to a 30-people jury who were asked to report which emotion was transmitted by the images presented in an on-line questionnaire, one image per page. The jury had free time to judge them and move to the next image. The musical

TABLE 2: Feature values adapted from Jongh [13].

| Feature | Image | Numerical |
|---------------------|---|---|
| <i>ieb_height</i> |  | $((a19.y + a21.y)/2 + (a20.y + a22.y)/2)/2$ |
| <i>oeb_height</i> |  | $(a17.y + a18.y)/2$ |
| <i>eb_frowned</i> |  | $((((a19.y - a21.y) + (a17.y - a21.y)) + ((a20.y - a22.y) + (a18.y - a22.y))))/2$ |
| <i>eb_slanting</i> |  | $((((a17.y - a21.y) + (a21.y - a19.y)) + ((a18.y - a22.y) + (a22.y - a20.y))))/2$ |
| <i>eb_distance</i> |  | $a20.x - a19.x$ |
| <i>eeb_distance</i> |  | $((a17.y - a7.y) + (a18.y - a8.y))/2$ |
| <i>e_openness</i> |  | $((a7.y - a5.y) + (a8.y - a6.y))/2$ |
| <i>e_slanting</i> |  | $((a3.y - a1.y) + (a4.y - a2.y))/2$ |
| <i>m_openness</i> |  | $\left(\frac{(-(a25.y - a26.y) - (a27.y - a29.y))}{2} \right) + \left(\frac{(-(a25.y - a26.y) - (a28.y - a30.y))}{2} \right) / 2$ |
| <i>m_mos</i> |  | $((a23.y - (a26.y + (\frac{a25.y - a26.y}{2}))) + (a24.y - (a26.y + (\frac{a25.y - a26.y}{2}))))/2$ |
| <i>m_width</i> |  | $a24.x - a23.x$ |
| <i>mul_height</i> |  | $(a29.y + a26.y + a30.y)/3$ |
| <i>mll_height</i> |  | $(a27.y + a25.y + a28.y)/3$ |
| <i>lc_height</i> |  | $(a23.y + a24.y)/2$ |

stimuli used in this study were not evaluated, since they had been evaluated in previous studies. The available alternatives for the 41 images (11 for *happiness*, 14 for *sadness* and 16 for *fear*) were *happiness*, *sadness*, *fear* or *none*. The results were evaluated to find consistency of judgements about the emotions provoked by the images, thereby supporting the selected visuals used for eliciting emotions. Figure 5 shows the ratings given by the jury.

For the purpose of describing the agreement among the user's reports, Fleiss' kappa (κ) values were obtained [79]. For *happiness* it was obtained $\kappa_h = 0.68$, which

indicates a substantial agreement, *sadness* and *fear* had a moderate agreement ($\kappa_s = 0.56$ and $\kappa_f = 0.52$) and *none* had a slight agreement ($\kappa_n = 0.03$), according to the interpretation of κ by Viera et al. [79]. The general κ was 0.45, which is considered moderate and therefore it implies in moderate agreement about the emotions elicited by the selected visual stimuli.

TABLE 3: AUs inference - criteria.

| AU | Criteria |
|----|---|
| 1 | $(ieb_height > (ieb_height_n + 0.075))$ and $(a20.y \geq a20_n.y)$ and $(a19.y \geq a19_n.y)$ |
| 2 | $(oeb_height > (oeb_height_n + 0.08))$ and $(a17.y \geq a17_n.y)$ and $(a18.y \geq a18_n.y)$ |
| 4 | $(ieb_height < (ieb_height - 0.03))$ and $(eb_distance < (eb_distance_n - 0.03))$ |
| 5 | $(e_slanting \geq (e_slanting_n - 0.05))$ and $(e_openness > (e_openness_n + 0.055))$ |
| 6 | $(m_mos \geq (m_mos_n + 0.045))$ and $(e_openness < (e_openness_n - 0.05))$ |
| 7 | $e_openness < (e_openness_n - 0.07)$ |
| 9 | $(m_width < (m_width_n - 0.1))$ and $(e_openness < (e_openness_n - 0.05))$ and $(ieb_height < (ieb_height_n - 0.06))$ |
| 10 | $(mul_height > (mul_height_n + 0.03))$ and $(m_width \leq m_width_n)$ and $(m_openness > (m_openness_n + 0.15))$ |
| 12 | $(m_mos \geq (m_mos_n + 0.05))$ and $(m_width > (m_width_n + 0.12))$ |
| 15 | $((m_mos + 0.03) \leq m_mos_n)$ and $(lc_height < (lc_height_n + 0.055))$ |
| 16 | $(mll_height \leq (mll_height_n - 0.1))$ and $(m_openness > (m_openness + 0.1))$ and $(lc_height < lc_height_n)$ |
| 17 | $(m_openness < m_openness_n)$ and $(abs(m_openness_n) - abs(m_openness)) \geq 0.08$ and $(-mll_height < (-mll_height_n - 0.08))$ |
| 20 | $(m_width > m_width_n)$ and $((abs(m_width) - abs(m_width_n)) \geq 0.15)$ and $(m_mos < (m_mos_n + 0.075))$ and $(lc_height < lc_height_n)$ |
| 23 | $m_openness < (m_openness_n - 0.1)$ |
| 24 | $(mul_height < mul_height_n)$ and $(mll_height > mll_height_n + 0.075)$ and $(m_openness < (m_openness_n - 0.1))$ |
| 25 | $m_openness \geq (m_openness_n + 0.13)$ |
| 26 | $(m_openness \geq (m_openness_n + 0.55))$ and $(m_openness \leq (m_openness_n + 0.63))$ |
| 27 | $m_openness \geq (m_openness_n + 0.63)$ |

TABLE 4: Tracks used for eliciting emotions.

| Emotion | Track | Album | Min:Sec |
|---------|---------------------------------|------------------------|---------------|
| Sad | Ask Your Saint Who He is Killed | The English Patient | 00:00 - 01:10 |
| Happy | Strip the Willow | Shallow Grave | 00:00 - 00:57 |
| Fear | Mussorgsky | Night on Bald Mountain | 00:00 - 01:00 |



Fig. 4: Visuals intended for eliciting emotions.

4.3 Procedure

The procedure can be described as follows:

- 1) The participant is explained the protocol in each step of the experiment. The nature of the research is not revealed, so the results are not harmed by

unconscious bias behaviour.

- 2) The participant is warned about the filming procedure and in order to enable data acquisition, a term sheet is given to formalise the cession of the obtained images for research purposes.
- 3) The participant is asked to fill a questionnaire about their musical experience, current feelings, if he/she is musically eclectic and often gets emotional listening to music.
- 4) A picture is taken from the individual in order to have a reference neutral facial expression. For those using glasses, it was asked to remove them during

²Source: <http://www.logueria.com/blog/91-como-ter-clientes-mais-felizes>. Access date: 09 March 2014.

³Source: <http://www.pictures88.com/comments/sad/>. Access date: 09 March 2014.

⁴Source: <http://www.dirtyandthirty.com/power/manifestation-monday-fear-what-are-you-scared-of-let-it-go/>. Access date: 20 December 2014.

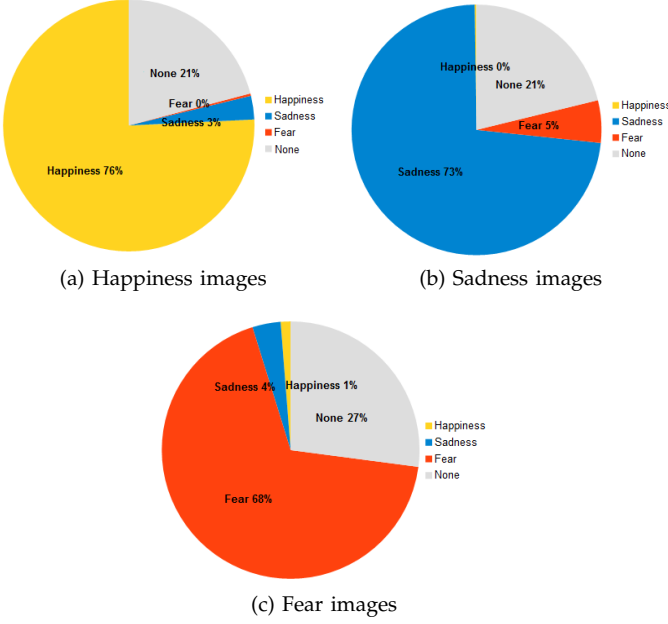


Fig. 5: Reports for visual stimuli regarding the emotions elicited by them.

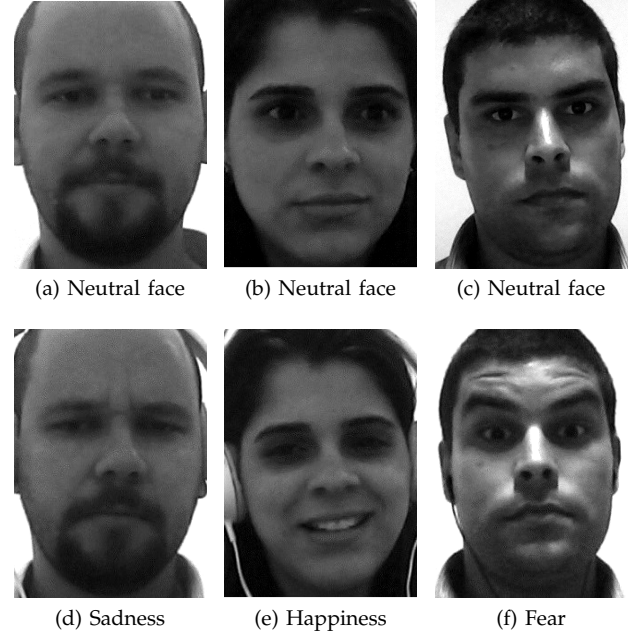


Fig. 6: Samples of expressive and neutral faces obtained during experiment.

the experiment.

- 5) At the same time, the participant hears a track and sees pictures displayed and changed every five seconds, both representing a specific emotion, while their face is being filmed. This process takes about 1 minute. At the end, an image is displayed warning the participant about the end of the experiment and then, the individual is notified about a second questionnaire.
- 6) The participant is asked to fill the second questionnaire about their new emotional state and a previous contact with the heard song. The latter question was necessary because, in some cases, it is easier to show emotional signals in the first contact with a song, since emotions, elicited by music, related to violation of expectancy might include anxiety/fear [57].

Facing the eliciting stimulus, participants presented facial expressions and the most perceptive ones were obtained from the *happiness* stimulus, as presented in other previous studies [80], [81]. Figure 6 shows samples collected during the experiment of still neutral faces – (a), (b) and (c) – and samples from the same users, but after they have been exposed to stimuli of *sadness*, *happiness* or *fear* – (d), (e) and (f) – respectively.

5 RESULTS AND DISCUSSION

5.1 Face action units recognition

In this study, images were obtained from 101 subjects who participated in the experiment presented in Section 4, considering only individuals without facial deformities, 52 females and 49 males, aged 18-60, approximately

33 per emotion. Approximately 270 frames (640×480 pixels) were obtained from each participant, plus an additional frame representing a neutral expression of each individual. Among the collected frames, the seven more expressive ones were selected for facial expression analysis. Forty images were used during training step in attempting to obtain thresholds for each rule of the system (see Table 3). After the training step, the 61 remaining images were submitted to the analyser. Table 5 contains the results from a set of about 700 images (training images + new images). Neutral expression occurrences were not considered in the analysis, unless when they caused false positives or false negatives.

The first column enumerates the AUs explored in this work according to the categorisation proposed by Ekman and Friesen [65]. The second column shows the occurrences of each AU, where only one occurrence was considered per subject (it varied between 0 and 1) without considering the frequency of AUs presented in frames of the same individual. The third column contains the AUs correctly classified by the analyser, taking as ground truth the human assessment. Columns *I*, *D* and *S* represent insertions, deletions, and substitutions, respectively. Insertions represent false positives which occur when an AU is found, but the ground truth states that the expression was neutral. For instance, AU 4 was found in three expressions considered neutral, therefore 4 insertions were obtained. On the other hand, deletions represent false negatives which occur when an AU is not found, but according to the ground truth the expression was observed in the image. For instance, an occurrence of AU 7 was not found by the analyser which inferred a neutral expression instead. Substitutions represent the

TABLE 5: AU recognition results on 101 subjects.

I: Insertions, D: Deletions, and S: Substitutions.

| AU | Occurrences | Correct | I | D | S | Accuracy | Precision |
|----|-------------|---------|---|---|---|----------|-----------|
| 1 | 11 | 11 | - | - | - | 100% | 100% |
| 2 | 11 | 11 | - | - | - | 100% | 100% |
| 4 | 16 | 16 | 3 | - | - | 100% | 81% |
| 5 | 4 | 4 | - | - | - | 100% | 100% |
| 6 | 18 | 18 | - | - | - | 100% | 100% |
| 7 | 51 | 50 | - | 1 | - | 98% | 96% |
| 9 | - | - | - | - | - | - | - |
| 10 | - | - | - | - | - | - | - |
| 12 | 27 | 26 | - | - | 1 | 96% | 93% |
| 15 | 15 | 14 | - | - | 1 | 93% | 87% |
| 16 | 4 | 4 | - | - | - | 100% | 100% |
| 17 | 5 | 5 | - | - | - | 100% | 100% |
| 20 | 2 | 2 | 2 | - | - | 100% | 0% |
| 23 | 10 | 10 | - | - | - | 100% | 100% |
| 24 | 1 | 1 | - | - | - | 100% | 100% |
| 25 | 17 | 17 | - | - | - | 100% | 100% |
| 26 | - | - | - | - | - | - | - |
| 27 | - | - | - | - | - | - | - |

occurrence of an AU X which was misclassified as an AU Y. For instance, AU 12 and AU 15 were misclassified as AU 20, hence there are 2 substitutions for them and 2 insertions for the latter. Finally, the columns accuracy and precision represent efficiency metrics grouped by AU. Both metrics were obtained by means of the following equations:

$$\begin{aligned} \text{accuracy} &= \frac{\text{correct}}{\text{occurrences}}, \\ \text{precision} &= \frac{\text{correct} - \text{incorrect}}{\text{occurrences}}. \end{aligned} \quad (1)$$

The accuracy describes the overall success rate, where the value for the system was 96%. On the other hand, the precision is a reliability metric which eliminates the correct results obtained due to chance. To do so, the number of correctly classified instances is subtracted from the misclassified ones (insertions + deletions + substitutions). The overall precision for the analyser was 92%.

5.2 Data analysis

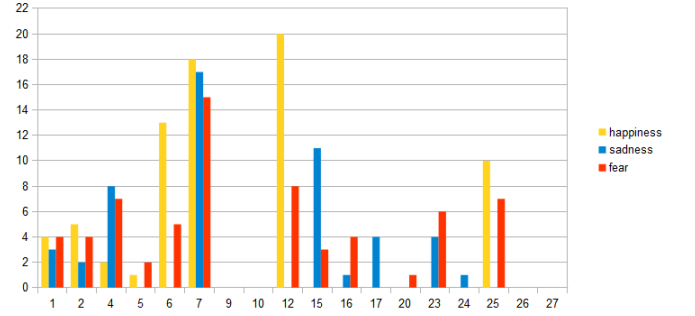
We have analysed images of 101 subjects submitted to stimuli intended for inducing emotional states and, consequently, unconscious facial responses, which are the focus of this study. Figure 7 shows some of the subjects who participated in the study.

Assuming the selected stimuli was able to evoke the expected emotions, it was possible to make a description of each emotion in terms of facial reactions using AUs. Figure 8 shows the histogram containing the presented AUs in each emotional scenario. The occurrences of

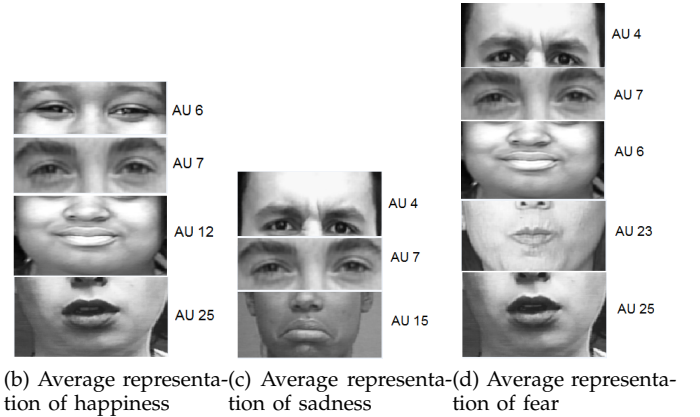


Fig. 7: Fifteen of the 101 participants in the experiment.

each AU were obtained without considering multiple occurrences for the same individual. For example, if an individual presented the AU 12 several times, only one occurrence of this AU would be considered for the analysis. The extraction of AUs was performed by the developed facial expression analyser jointly with human assessment in order to ensure the accuracy of the results.



(a) Histogram of AUs



(b) Average representation of happiness (c) Average representation of sadness (d) Average representation of fear

Fig. 8: AUs grouped by emotion.

In *Happiness*, it was possible to find a pattern in facial reactions presented by the individuals. The majority of them presented the AUs 6, 7, 12, and, in some situa-

tions, AU 25. The pattern found is consistent with the prototypic representation of this emotion described in the literature, differing only regarding the occurrence of AU 7, which is associated with lowering the eyelids during spontaneous smiles, but it is rarely presented in artificial smiles. In *Sadness*, the predominant AUs were 7, 15 and 4, where AU 15 had an extremely smooth representation, in contrast to the prototypic representation for this emotion. AUs 4 and 7 were presented signalling discomfort caused by the eliciting stimulus. Apathy was also observed during the exposure to the sadness stimulus. The most significant difference was found in *Fear*. The predominant AUs were 7, 12, 4, 23 e 25 (either single or combined displays), where AU 7 was presented by individuals closing their eyes during image exposure, and AU 12 was presented in smooth smiles. Surprisingly, the smiles presented during *Fear* have scientific explanation. Allan and Barbara Pease [82] presented 2 types of smiles, being one of them known as a fear face, performed by primates. This fear smile communicates submission or anxiety facing a fear scenario, which is why some people smile thinking about a frightening situation, it is an involuntary way of protecting themselves by showing submission to the threat. AU 23 is an indicative of anxiety as well, once this action is associated with the dryness of mouth caused by situations of anxiety related to fear, for example. AU 25 was observed with some frequency in some individuals in a smooth representation, different from the prototypic exhibition of the emotion fear. The most expressive expressions were found in *Happiness*, other scenarios presented low intensity expressions, or even neutral ones.

The data obtained through questionnaires and the employed emotion stimuli were compared to the appraisal reported by the individuals aiming to identify whether particular parameters affect emotional experiences.

We accounted for the occurrences of each given question and compared them to the reported appraisal in order to obtain the correlation measure between them. In addition, we compared the appraisal to the emotional stimulus to check its validity. The purpose of this study is not to provide a formula for emotion inference, rather it is to highlight the impact of the proposed dimensions on the final emotional state, statistically, which implies that they should not be neglected in emotion assessment scenarios.

The evaluated parameters were the emotion stimulus, mood, musical experience, subject's preferences, emotionality, previous contact with the musical stimulus, gender, and age. Figure 9 (parameters at the same scale) and Figure 10 (parameters at different scales) show the scatter plots containing the data distribution according to its occurrence in the analysed sample.

The points represent the combination of appraisal and explored measure, which are shown in a gray scale where the darkest points correspond to highest occurrences in the sample, which happens when several

subjects describe the same values for the emotional parameters, and the lighter ones correspond to lowest occurrences. The red line represents a linear trend line⁵ which shows how data is expected to be scattered in the sample considering its distribution. The emotional parameters were described by numeric values, where the stimulus is represented by 1 as *Happiness*, 2 as *Fear*, and 3 as *Sadness* (see Figure 10a), Appraisal and Mood are represented in a 1-5 scale, 5 being the highest value (see Figure 9a), Known song is represented in a 0-1 scale, 0 being the absence of previous contact with the song played during the experiment and 1 being the existence of this previous contact (see Figure 10b), Gender is represented by 0 as *Male* and 1 as *Female* (see Figure 10c) and Age is represented by 0 as ≤ 25 , 1 as 25-48 and 2 as 48-60 (see Figure 10d).

Pearson's correlation coefficient (ρ) [83] was obtained to quantify the correlation between the variables presented in Figure 9, where it was found moderate correlation between the emotion stimulus and appraisal ($\rho_{s,a} \approx -0.62$) and weak correlation between other variables⁶. These results demonstrate that the lower the stimulus (in terms of the numeric labels for each emotion, as previously explained) the greater the reported appraisal, which suggests that the emotion stimuli were able to elicit emotional experiences since, in most cases, the reported appraisal was consistent with the expected emotion in each emotional scenario.

The analysis of emotional parameters also considered each emotion individually. The obtained results indicated moderate correlation between mood and appraisal for *Happiness* and *Fear*, where the correlation coefficient obtained was $\rho \approx 0.58$ for happiness stimulus (see Figure 11a) and $\rho \approx 0.6$ for fear stimulus (see Figure 11b), which means that the higher the mood, the higher the reported appraisal in these emotional scenarios. For sadness stimulus the correlation was weak (see Figure 11c), which demonstrates that for this emotion the mood did not affect the emotional experience (in terms of reported appraisal). Furthermore, it was found a moderate correlation between appraisal and musical preferences for *Fear* ($\rho \approx 0.42$) indicating that in this emotional scenario the subjects who reported higher appraisal were the more musically eclectic ones (see Figure 11d).

Considering a gender distinction in the sample, it was found a moderate correlation ($\rho \approx 0.54$) between *Mood* and *Appraisal* for males (see Figure 12a), while for females, the mood did not affect their emotional experience in most individuals (see Figure 12b). These results suggest that the higher the mood, the higher the appraisal reported by males, thus supporting the superior performance of females over males in emotion analysis since the mood did not affect females'

⁵Source: <https://support.office.com/en-US/Article/Add-a-trend-or-moving-average-line-to-a-chart-3c4323b1-e377-43b9-b54b-fae160d97965>. Access date: 18 March 2016.

⁶The terms used to describe the strength of the correlation were obtained from Gerstman [83].

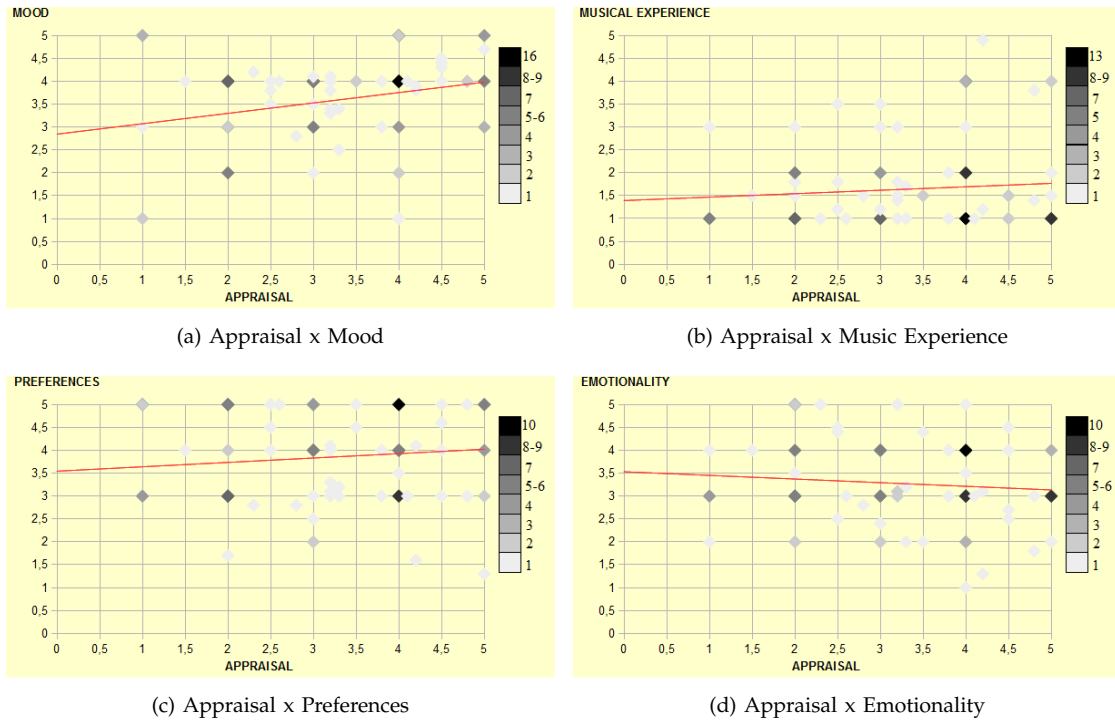


Fig. 9: Scatter plot from emotional parameters: mood, music experience, preferences, and emotionality.

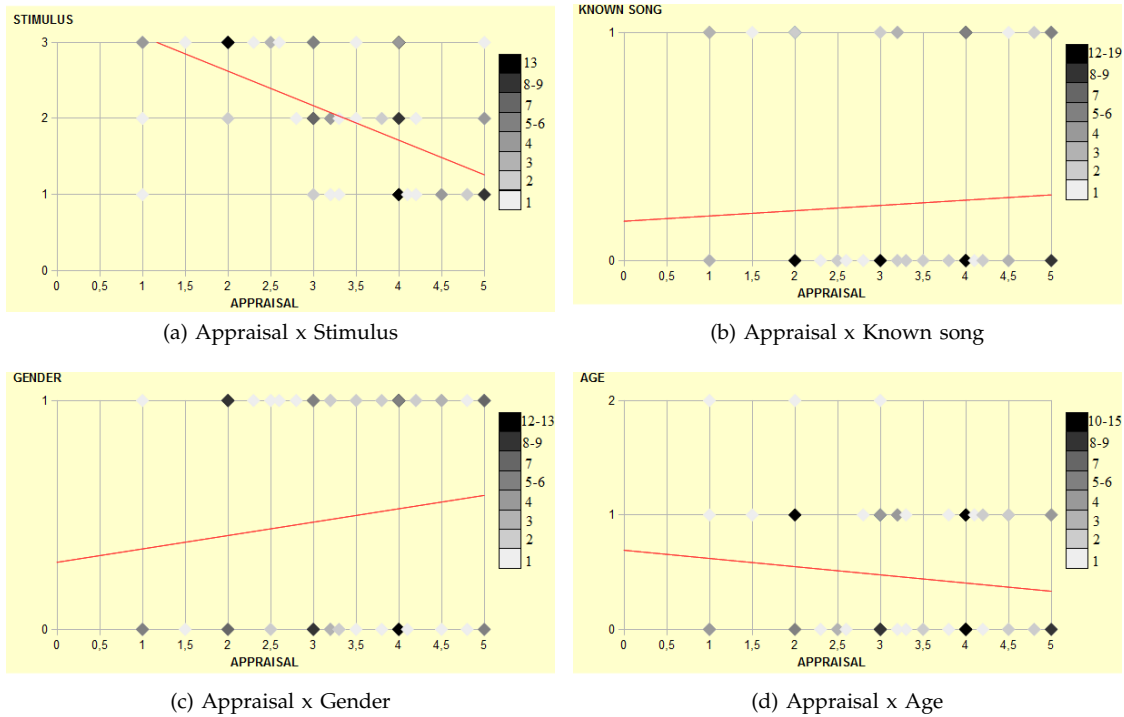


Fig. 10: Scatter plot from emotional parameters: stimulus, known song, gender, and age.

performance in this task. Concerning the correlation between *Appraisal* and *Stimulus*, both males and females presented moderate values ($\rho \approx -0.62$), which means that the eliciting stimulus provoked the same emotional states regardless subject's gender (see Figure 12 - (c) and

(d)).

In summary, the overall results suggested that the designed stimuli were able to elicit emotions since more than 65% of the users reported changes in mood (Figure 9a). In addition, the results showed the effect of mood on

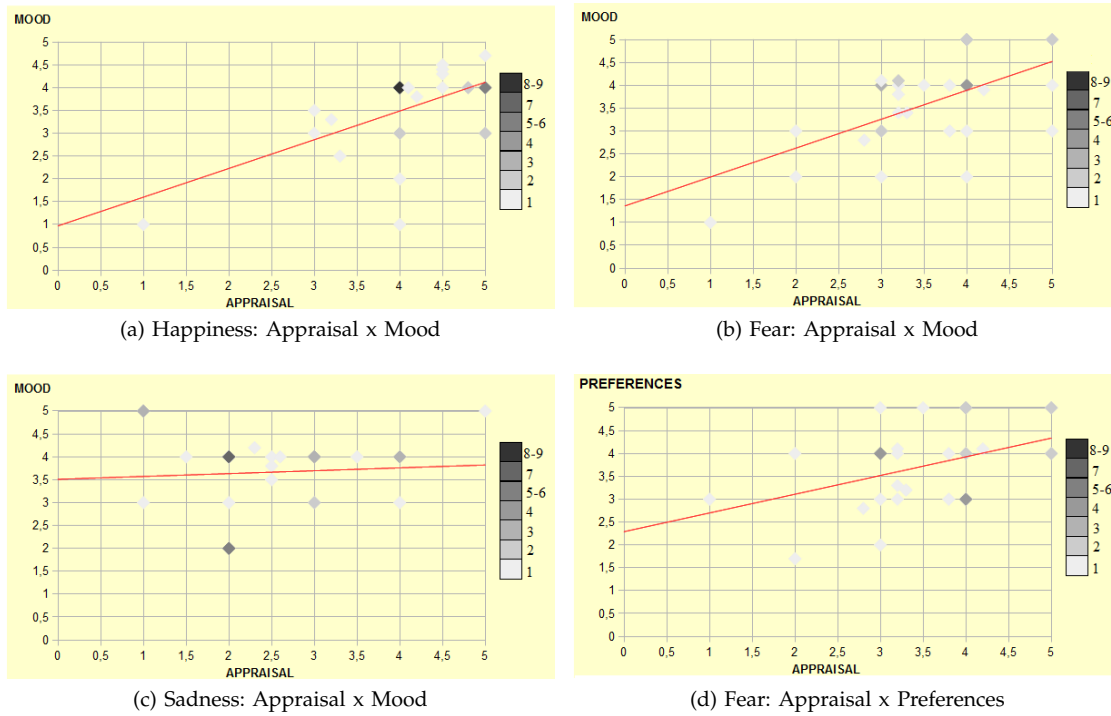


Fig. 11: Scatter plot from emotional parameters grouped by emotion.

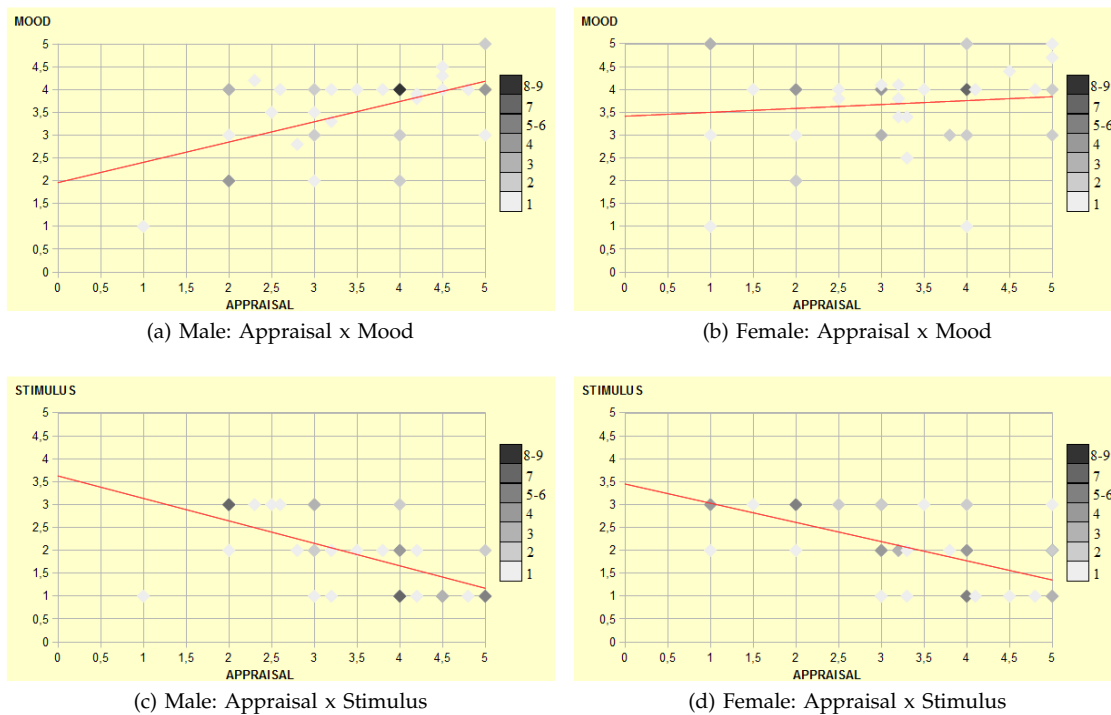


Fig. 12: Scatter plot from emotional parameters grouped by gender.

emotional experiences of happiness and fear and the influence of musical preferences on the appraisal reported by individuals who had contact with the fear stimulus. The analysis of emotional parameters suggests that the four dimensions were valid to describe the emotional

experience, although some subjective factors could have been disregarded (e.g. music experience, emotionality, and known song). Finally, 79% of the users reported an appraisal consistent with the used stimulus (Figure 10a).

We understand these are initial results; nevertheless,

we believe they point to very significant differences concerning emotion analysis from facial expressions, mainly due to the use of spontaneous facial expressions instead of prototypic ones. In the proposed experimental study, we have presented results for only three basic emotions, but we believe that this method can be used to other emotions.

6 FINAL CONSIDERATIONS

The presented framework brings a new perspective on facial expression analysis considering spontaneous emotional experiences in order to find a system which is most suitable for human interaction studies. We have demonstrated that there is a link between subjective information and experienced emotions, which encourages the use of such information in emotion analysis studies. We also hope to encourage the use of more realistic data in the area of facial expression analysis by providing a method to elicit spontaneous emotional experiences. For future work, we will cover all four basic emotions by creating a visual-sonorous stimulus for inducing natural facial expressions for each emotion, thereby improving the collected set of images. Additionally, the age range will be expanded to contain younger and older individuals, which will allow a more robust data analysis taking into account distinct age groups. Finally, the database will be publicly released to support the development and assessment of further facial expression analysers.

REFERENCES

- [1] C. Krumhansl, "An exploratory study of musical emotions and psychophysiology," *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, vol. 51, no. 4, pp. 336–353, 1997.
- [2] R. Picard, E. Vyzas, and J. Healey, "Toward machine emotional intelligence: Analysis of affective physiological state," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 10, pp. 1175–1191, 2001.
- [3] R. Adolphs, "Recognizing emotion from facial expressions: Psychological and neurological mechanisms," *Behavioral and cognitive neuroscience reviews*, vol. 1, no. 1, pp. 21–62, 2002.
- [4] K. Kim, S. Bang, and S. Kim, "Emotion recognition system using short-term monitoring of physiological signals," *Medical and Biological Engineering and Computing*, vol. 42, no. 3, pp. 419–427, 2004.
- [5] G. Chanel, J. Kronegg, D. Grandjean, and T. Pun, "Emotion assessment: Arousal evaluation using EEG's and peripheral physiological signals," in *Multimedia Content Representation, Classification and Security*, ser. Lecture Notes in Computer Science, B. Gunsel, A. Jain, A. Tekalp, and B. Sankur, Eds. Springer Berlin Heidelberg, 2006, vol. 4105, pp. 530–537.
- [6] S. Le Groux, A. Valjamae, J. Manzolli, and P. F. Verschure, "Implicit physiological interaction for the generation of affective musical sounds," *Proceedings of the International Computer Music Conference*, 2008.
- [7] A. F. Smeaton and S. Rothwell, "Biometric responses to music-rich segments in films: The cdvplex," in *International workshop on content-based multimedia indexing*, 2009, pp. 162–168.
- [8] T. Colibazzi, J. Posner, Z. Wang, D. Gorman, A. Gerber, S. Yu, H. Zhu, A. Kangarlu, Y. Duan, J. Russell, and B. Peterson, "Neural systems subserving valence and arousal during the experience of induced emotions," *Emotion*, vol. 10, no. 3, pp. 377–389, 2010.
- [9] N. Dael, M. Mortillaro, and K. R. Scherer, "Emotion expression in body action and posture," *Emotion*, vol. 12, no. 5, pp. 1085–1101, 2012.
- [10] L. Jiang, Z. Qing, and W. Wenyuan, "A novel approach to analyze the result of polygraph," in *Proceedings of IEEE International Conference on Systems, Man and Cybernetics*, vol. 4, 2000, pp. 2884–2886.
- [11] H. Kobayashi and F. Hara, "The recognition of basic facial expressions by neural network," in *Proceedings to the IEEE International Joint Conference on Neural Networks*, vol. 1, 1991, pp. 460–466.
- [12] Y. Tian, T. Kanade, and J. Cohn, "Recognizing action units for facial expression analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 2, pp. 97–115, 2001.
- [13] E. Jongh, "Fed: An online facial expression dictionary as a first step in the creation of a complete nonverbal dictionary," Master's thesis, Delft University of Technology, June 2002.
- [14] I. Cohen, N. Sebe, F. Gozman, M. C. Cirelo, and T. S. Huang, "Learning bayesian network classifiers for facial expression recognition both labeled and unlabeled data," in *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*, vol. 1, 2003, pp. 1–595.
- [15] C. Chuang and F. Shih, "Recognizing facial action units using independent component analysis and support vector machine," *Pattern Recognition*, vol. 39, no. 9, pp. 1795–1798, 2006.
- [16] L. Veloso, J. Carvalho, C. Cavalcanti, E. Moura, F. Coutinho, and H. Gomes, "Neural network classification of photogenic facial expressions based on fiducial points and gabor features," in *Advances in Image and Video Technology*, ser. Lecture Notes in Computer Science, D. Mery and L. Rueda, Eds., 2007, vol. 4872, pp. 166–179.
- [17] P. Yang, Q. Liu, and D. N. Metaxas, "Boosting coded dynamic features for facial action units and facial expression recognition," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2007, pp. 1–6.
- [18] J. Hamm, C. Kohler, R. Gur, and R. Verma, "Automated facial action coding system for dynamic analysis of facial expressions in neuropsychiatric disorders," *Journal of Neuroscience Methods*, vol. 200, no. 2, pp. 237–256, 2011.
- [19] W. Quan, B. Matuszewski, L. Shark, and C. Frowd, "Methodology and performance analysis of 3-D facial expression recognition using statistical shape representation," *International Journal of Grid and Distributed Computing*, vol. 4, no. 3, pp. 79–88, 2011.
- [20] A. Savran, B. Sankur, and M. Taha Bilge, "Regression-based intensity estimation of facial action units," *Image and Vision Computing*, vol. 30, no. 10, pp. 774–784, 2012.
- [21] M. Valstar and M. Pantic, "Fully automatic recognition of the temporal phases of facial actions," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 42, no. 1, pp. 28–43, 2012.
- [22] F. Dornaika, A. Moujahid, and B. Raducanu, "Facial expression recognition using tracked facial actions: Classifier performance analysis," *Engineering Applications of Artificial Intelligence*, vol. 26, no. 1, pp. 467–477, 2013.
- [23] E. Sariyanidi, H. Gunes, and A. Cavallaro, "Automatic analysis of facial affect: A survey of registration, representation and recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 99, no. PrePrints, p. 1, 2014.
- [24] F. De la Torre and J. Cohn, "Facial expression analysis," in *Visual Analysis of Humans*, T. B. Moeslund, A. Hilton, V. Krüger, and L. Sigal, Eds., 2011, pp. 377–409.
- [25] P. Ekman, W. V. Friesen, M. O'Sullivan, A. Chan, I. Diacyanni-Tarlatzis, K. Heider, R. Krause, W. A. LeCompte, T. Pitcairn, P. E. Ricci-Bitti et al., "Universals and cultural differences in the judgments of facial expressions of emotion," *Journal of personality and social psychology*, vol. 53, no. 4, pp. 712–717, 1987.
- [26] R. E. Jack, O. G. Garrod, and P. G. Schyns, "Dynamic facial expressions of emotion transmit an evolving hierarchy of signals over time," *Current Biology*, vol. 24, no. 2, pp. 187–192, 2014.
- [27] J. F. Cohn and K. L. Schmidt, "The timing of facial motion in posed and spontaneous smiles," *J. Wavelets, Multi-resolution & Information Processing*, vol. 2, pp. 1–12, 2004.
- [28] M. Bartlett, G. Littlewort, M. Frank, C. Lainscsek, I. Fasel, and J. Movellan, "Recognizing facial expression: machine learning and application to spontaneous behavior," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2005. CVPR 2005., vol. 2, 2005, pp. 568–573.
- [29] M. F. Valstar, M. Pantic, Z. Ambadar, and J. F. Cohn, "Spontaneous vs. posed facial behavior: Automatic analysis of brow actions," in *Proc. ACM Int'l Conf. Multimodal Interfaces*, 2006, pp. 162–170.

- [30] Z. Zeng, Y. Fu, G. I. Roisman, Z. Wen, Y. Hu, and T. S. Huang, "Spontaneous emotional facial expression detection," *Journal of Multimedia*, vol. 1, no. 5, pp. 1–8, 2006.
- [31] S. Lucey, A. B. Ashraf, and J. Cohn, "Investigating spontaneous facial action recognition through aam representations of the face," in *Face Recognition Book*, K. Kurihara, Ed., 2007, pp. 275–286.
- [32] Z. Zeng, M. Pantic, G. Roisman, and T. Huang, "A survey of affect recognition methods: Audio, visual, and spontaneous expressions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 1, pp. 39–58, 2009.
- [33] H. Gunes and B. "Categorical and dimensional affect analysis in continuous input: Current trends and future directions," *Image and Vision Computing*, vol. 31, no. 2, pp. 120–136, 2013, affect Analysis In Continuous Input.
- [34] P. Ekman and W. Friesen, "Measuring facial movement," *Environmental psychology and nonverbal behavior*, vol. 1, no. 1, pp. 56–75, 1976.
- [35] J. Bazzo and M. Lamar, "Recognizing facial actions using gabor wavelets with neutral face average difference," in *Proceedings of IEEE International Conference on Automatic Face and Gesture Recognition*, May 2004, pp. 505–510.
- [36] D. Vukadinovic and M. Pantic, "Fully automatic facial feature point detection using gabor feature based boosted classifiers," in *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 2, 2005, pp. 1692–1698.
- [37] C. Deruelle, C. Rondan, B. Gepner, and C. Tardif, "Spatial frequency and face processing in children with autism and Asperger syndrome," *Journal of autism and developmental disorders*, vol. 34, no. 2, pp. 199–210, 2004.
- [38] M. Robin, A. Pham-Scottez, F. Curt, C. Dugre-Le Bigre, M. Speranza, D. Sapinho, M. Corcos, S. Berthoz, and G. Kedia, "Decreased sensitivity to facial emotions in adolescents with borderline personality disorder," *Psychiatry research*, vol. 200, no. 2, pp. 417–421, 2012.
- [39] A. R. Daros, K. K. Zakzanis, and A. C. Ruocco, "Facial emotion recognition in borderline personality disorder," *Psychological Medicine*, vol. 43, pp. 1953–1963, 2013.
- [40] J. M. Girard, J. F. Cohn, M. H. Mahoor, S. Mavadati, and D. P. Rosenwald, "Social risk and depression: Evidence from manual and automatic facial expression analysis," in *Proceedings of IEEE International Conference on Automatic Face and Gesture Recognition*, 2013, pp. 1–8.
- [41] C. J. Duthoit, T. Sztynka, S. K. L. Lal, B. T. Jap, and J. I. Agbinya, "Optical flow image analysis of facial expressions of human emotion: Forensic applications," in *Proceedings of the 1st International Conference on Forensic Applications and Techniques in Telecommunications, Information, and Multimedia and Workshop, ser. e-Forensics '08*. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2008, pp. 5:1–5:6.
- [42] M. Wimmer, B. MacDonald, D. Jayamuni, and A. Yadav, "Facial expression recognition for human-robot interaction - a prototype," in *Robot Vision*, ser. Lecture Notes in Computer Science, G. Sommer and R. Klette, Eds., 2008, vol. 4931, pp. 139–152.
- [43] T. Nakanishi and T. Kitagawa, "Visualization of music impression in facial expression to represent emotion," in *Proceedings of the Asia-Pacific Conference on Conceptual Modelling*, vol. 53, 2006, pp. 55–64.
- [44] S. Sun and C. Ge, "A new method of 3D facial expression animation," *Journal of Applied Mathematics*, vol. 2014, pp. 1–6, 2014.
- [45] C. Laurier, J. Grivolla, and P. Herrera, "Multimodal music mood classification using audio and lyrics," in *Proceedings of the International Conference on Machine Learning and Applications*, San Diego, California (USA), December 2008, pp. 688–693.
- [46] P. Tyler, "Developing a two-dimensional continuous response space for emotions perceived in music," Ph.D. dissertation, Florida State University, 1996.
- [47] E. Schubert, "Continuous response to music using a two dimensional emotion space," in *Proceedings of the International Conference of Music Perception and Cognition*, 1996, pp. 263–268.
- [48] J. Fontaine, K. Scherer, E. Roesch, and P. Ellsworth, "The world of emotions is not two-dimensional," *Psychological Science*, vol. 18, no. 12, pp. 1050–1057, 2007.
- [49] M. Trkulja and D. Janković, "Towards three-dimensional model of affective experience of music," in *Emotion*, vol. 17, 2012, pp. 25–40.
- [50] U. Schimmack and A. Grob, "Dimensional models of core affect: A quantitative comparison by means of structural equation modeling," *European Journal of Personality*, vol. 14, no. 4, pp. 325–345, 2000.
- [51] J. D. Morris, N. J. Klahr, F. Shen, J. Villegas, P. Wright, G. He, and Y. Liu, "Mapping a multidimensional emotion in response to television commercials," *Human Brain Mapping*, vol. 30, no. 3, pp. 789–796, 2009.
- [52] K. R. Scherer, "The dynamic architecture of emotion: Evidence for the component process model," *Cognition & Emotion*, vol. 23, no. 7, pp. 1309–1316, 2009.
- [53] J. K. Vuoskoski and T. Eerola, "Measuring music-induced emotion a comparison of emotion models, personality biases, and intensity of experiences," *Musicae Scientiae*, vol. 15, no. 2, pp. 159–173, 2011.
- [54] P. C. Ellsworth and K. R. Scherer, "Appraisal processes in emotion," *Handbook of affective sciences*, vol. 572, pp. 572–595, 2003.
- [55] K. R. Scherer, "What are emotions? and how can they be measured?" *Social science information*, vol. 44, no. 4, pp. 695–729, 2005.
- [56] H. G. Wallbott and K. R. Scherer, "Assessing emotion by questionnaire," *Emotion: Theory, research, and experience*, vol. 4, pp. 55–82, 1989.
- [57] P. Juslin, "From everyday emotions to aesthetic emotions: Towards a unified theory of musical emotions," *Physics of Life Reviews*, vol. 10, no. 3, pp. 235–266, 2013.
- [58] A. Mehrabian, *Communication without words*. Psychological today, 1968, vol. 2, no. 4.
- [59] J. Posner, J. Russell, and B. Peterson, "The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology," *Development and Psychopathology*, vol. 17, no. 3, pp. 715–734, 2005.
- [60] Y.-L. Tian, T. Kanade, and J. F. Cohn, "Facial expression analysis," in *Handbook of face recognition*, 2005, ch. 11, pp. 247–275.
- [61] C. Darwin, *The expression of the emotions in man and animals*, 3rd ed. Oxford University Press, 1998.
- [62] H. Hoffmann, H. Kessler, T. Eppel, S. Rukavina, and H. C. Traue, "Expression intensity, gender and facial emotion recognition: Women recognize only subtle facial emotions better than men," *Acta Psychologica*, vol. 135, no. 3, pp. 278–283, 2010.
- [63] U.-S. Donges, A. Kersting, and T. Suslow, "Women's greater ability to perceive happy facial emotion automatically: gender differences in affective priming," *PloS one*, vol. 7, no. 7, p. e41745, 2012.
- [64] R. Nauert, "Women recognize emotions better," *Psych Central*, 2009, access date: 11 October 2014. [Online]. Available: <http://psychcentral.com/news/2009/10/22/women-recognize-emotions-better/9100.html>
- [65] P. Ekman, W. Friesen, and J. Hager, *Facial Action Coding System (FACS): Manual*. A Human Face, 2002.
- [66] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1, 2001, pp. 511–518.
- [67] R. Gonzalez and R. Woods, *Digital Image Processing*, 3rd ed. Pearson/Prentice Hall, 2008.
- [68] G. Bradski and A. Kaehler, *Learning OpenCV: Computer vision with the OpenCV library*, 2nd ed., M. Loukides, Ed. O'Reilly Media, Inc., September 2008.
- [69] E. Douglas-Cowie, N. Campbell, R. Cowie, and P. Roach, "Emotional speech: Towards a new generation of databases," *Speech communication*, vol. 40, no. 1, pp. 33–60, 2003.
- [70] A. J. O'Toole, J. Harms, S. L. Snow, D. R. Hurst, M. R. Pappas, J. H. Ayyad, and H. Abdi, "A video database of moving faces and people," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 5, pp. 812–816, 2005.
- [71] M. Pantic, M. Valstar, R. Rademaker, and L. Maat, "Web-based database for facial expression analysis," in *IEEE International Conference on Multimedia and Expo, ICME 2005*, 2005, pp. 317–321.
- [72] J. Sloboda and P. Juslin, "Psychological perspectives on music and emotion," in *Music and emotion: Theory and research*, 2001, pp. 71–104.
- [73] X. Hu, "Music and mood: Where theory and reality meet," *Proceedings of iConference*, 2010.
- [74] M. Korsakova-Kreyn and W. J. Dowling, "Emotion in music: Affective responses to motion in tonal space," *Proceedings of the 12th International Conference on Music Perception and Cognition and*

the 8th Triennial Conference of the European Society for the Cognitive Sciences of Music, pp. 23–28, 2012.

- [75] A. Capurso and M. R. Foundation, *Music and your emotions: a practical guide to music selections associated with desired emotional responses*, L. P. Corp., Ed. Liveright Pub. Corp., 1952.
- [76] T. Eerola and J. Vuoskoski, "A comparison of the discrete and dimensional models of emotion in music," *Psychology of Music*, vol. 39, no. 1, pp. 18–49, 2010.
- [77] T. Gašpar, M. Labor, I. Jurić, D. Dumančić, V. Ilakovac, and M. Heffer, "Comparison of emotion recognition from facial expression and music," *Collegium antropologicum*, vol. 35, no. 1, pp. 163–167, 2011.
- [78] U. Hess, P. Philippot, and S. Blairy, "Facial reactions to emotional facial expressions: Affect or cognition?" *Cognition & Emotion*, vol. 12, no. 4, pp. 509–531, 1998.
- [79] A. J. Viera, J. M. Garrett *et al.*, "Understanding interobserver agreement: the kappa statistic," *Fam Med*, vol. 37, no. 5, pp. 360–363, 2005.
- [80] M. J. Black and Y. Yacoob, "Recognizing facial expressions in image sequences using local parameterized models of image motion," *International Journal of Computer Vision*, vol. 25, no. 1, pp. 23–48, 1997.
- [81] C. L.-C. Huang, S. Hsiao, H.-G. Hwu, and S.-L. Howng, "The chinese facial emotion recognition database (CFERD): A computer-generated 3-D paradigm to measure the recognition of facial emotional expressions at different intensities," *Psychiatry Research*, vol. 200, no. 2–3, pp. 928–932, 2012.
- [82] A. Pease and B. Pease, *The definitive book of body language*. Random House LLC, 2008.
- [83] B. B. Gerstman, "Statprimer," 2003, access date: 08 November 2014. [Online]. Available: <http://www.sjsu.edu/faculty/gerstman/StatPrimer/>