Int. J. Human-Computer Studies 83 (2015) 51–61

Contents lists available at ScienceDirect

Int. J. Human-Computer Studies

journal homepage: www.elsevier.com/locate/ijhcs

# *Oudjat*: A configurable and usable annotation tool for the study of facial expressions of emotion ☆

Damien Dupre [a,*], Daniel Akpan [b], Elena Elias [c], Jean-Michel Adam [b], Brigitte Meillon [b,d], Nicolas Bonnefond [b,e], Michel Dubois [a], Anna Tcherkassof [a]

[a] Univ. Grenoble Alpes, LIPPC2S, F-38000 Grenoble, France
[b] Univ. Grenoble Alpes, LIG, F-38000 Grenoble, France
[c] Floralis—UJF filiale, F-38610 Gières, France
[d] CNRS, LIG, F-38000 Grenoble, France
[e] INRIA, F-38330 Montbonnot-Saint-Martin, France

ARTICLE INFO

*Article history:*
Received 4 July 2014
Received in revised form
27 May 2015
Accepted 28 May 2015
Communicated by Nicu Sebe
Available online 26 June 2015

*Keywords:*
Annotation
Emotion
Facial expressions
Open-source software
Observation research

ABSTRACT

This paper describes *Oudjat*, a new software program which has been developed in order to conduct recognition experiments. *Oudjat* is dedicated to the manual annotation facial expressions of emotion (FEE). Considering the existence of other software applications in that field, *Oudjat* provides a compromise solution between the currently existing tools. For the investigators, it is an easy-to-configure interface to set up relevant behaviors to be annotated. For the annotators, it is an easy-to-use interface. This tool can perform complex annotations procedures utilizing multiple responses panels such as buttons, scales (e.g., Likert scales), and free labelling. *Oudjat* also allows to chain response panels or to conduct sequence marking annotations (i.e., two-steps temporal annotation). As it can be configured in any language, *Oudjat* is particularly suited for intercultural experiments. Four annotation procedures are presented to illustrate *Oudjat*'s possibilities with FEE annotation. *Oudjat* is an open source software available to the scientific community, and can be freely be obtained on request.



## 1. Introduction

Emotions are widespread social reactions in human societies. They can be defined as the result of spontaneous and quick internal and external modifications triggered by a stimulus (Tcherkassof, 2008). They are relevant cues not only for communicating feelings but also for regulating daily interactions through different channels such as voice or body posture (Lottridge et al., 2011). Among these channels, facial expressions of emotion (FEE) have long been, and still are, the main means to study the recognition of emotions since they are considered to be the privileged index of the experienced emotion.

FEE studies are performed using either objective methods such as measuring the activity of facial muscles by sensors (e.g., electromyography), or subjective ones such as the annotation procedures. Although objective, electromyography measurements are invasive and are therefore incompatible with the study of natural FEE. Not invasive, the annotation procedure is thus often preferred to study FEE. Two different annotation procedures could be distinguished: manual annotation and automatic annotation of FEE. The automatic annotation of FEE is based on the detection of facial features (such as the mouth, the eyes and eyebrows) and emotions are classified on the basis of these features (Sariyanidi et al., 2014). Automatic annotation is robust and has good recognition rates with prototypical FEE (Metaxas and Zhang, 2013). These displays are intense enough to be recognized as emotional expressions and to be differentially classified according to the displayed emotion. However, automatic annotation performance is significantly lowered for less prototypical FEE and/or for FEE combining various emotions. For the latters, manual annotation is more suited. It consists in having participants judging emotions displayed by faces. This procedure highlights how others' emotional displays are perceived and emotionally interpreted (Kelly and Agnew, 2011). Presenting a new tool to perform manual annotation experiments, the present paper focuses on the manual annotation of FEE stimuli. It will first detail the key elements for such annotation, because different variants of annotation procedures are distinguished according to the kind of stimuli studied (Section 2.1), the type of judgment required from the participant (Section 2.2) and the kind of annotation procedure carried out (Section 2.3). It will then present the existing major

☆This paper has been recommended for acceptance by Nicu Sebe.

* Corresponding author. Tel.: +33 4 76 82 58 98; fax: +33 4 76 82 56 65.
*E-mail address:* Damien.dupre@upmf-grenoble.fr (D. Dupre).

http://dx.doi.org/10.1016/j.ijhcs.2015.05.010
1071-5819/

tools for the manual annotation of FEE (Section 2). Finally, a new tool called *Oudjat* will be exposed, followed by the description of emotion recognition investigations using this software (Section 3).

## 2. Key elements for the study of FEE

### 2.1. FEE stimuli

In order to understand how facial expressions are recognized, stimuli displaying various persons' faces after an emotional elicitation are used by researchers. These recordings can either be natural or posed. *Natural* FEE stimuli are recorded when people express emotions spontaneously whereas *posed* FEE stimuli are deliberately portrayed either by actors or by lay persons. FEE stimuli can also be *static* (pictures) or *dynamic* (videos). Static stimuli are widely used for FEE annotation because they are easy to manipulate and to be used in experiments. It is the case either when functional Magnetic Resonance Imaging (fMRI) or Electroencephalography (EEG) studies are conducted. Such studies must include technical constraints that recommend the use of static FEE. For example, all stimuli of fMRI studies must display equated for mean luminance (i.e., the luminous intensity of the faces) to compare how each FEE is more or less accurately recognized, and it is quite easy to manipulate the luminance's level of static stimuli (e.g., Kim et al., 2003; or Winston et al., 2004). However, static stimuli do not convey the temporal modifications of the face which are important features for FEE recognition (Krumhuber et al., 2013). Indeed, static FEE are less informative than dynamic FEE which convey movements and motions of the face such as timing and regularity unfolding. For example, spontaneous FEE can change very quickly from a specific emotion to another. Stimuli can also be mixed or compound FEE displaying more than one emotion (Du et al., 2014; Sullivan and Strongman, 2003). Facial movements are especially important for the accurate recognition of natural FEE. Less intense than posed FEE, they benefit from temporal information. However, dynamic FEE are harder to handle than static ones. For example, contrary to pictures, it is difficult to increase or decrease the luminance of video stimuli (Zhao and Pietikainen, 2007). This is why, for instance, FEE videos cannot be used for fMRI studies. In sum, static and dynamic FEE are both relevant stimuli for the study of emotional recognition but they both have their own benefits and flaws. Static FEE are easy to use but they are not natural enough. Dynamic FEE are closer to natural facial expressions but they need complex procedures to be handled.

### 2.2. Type of facial emotional information

Whether natural or posed, FEE consist in facial muscle configurations that can be dynamic or static. In both cases, there are two different ways to account for their recognition: either in terms of sign judgment or in terms of message judgment (Pantic, 2009). Sign judgment is based on the identification of separate facial movements. The foremost taxonomy of facial muscle configurations is the Facial Action Coding System (FACS) in which each facial muscle movement is coded as an action unit or AU (Cohn et al., 2007; Ekman and Friesen, 1978). The FACS inventories each specific AU configurations associated to specific emotions. For example, the configuration of AU6, corresponding to cheek raiser muscle (*orbicularis oculi*), associated with AU12, corresponding to lip corner puller muscle (*zygomaticus major*), constitutes the prototypical facial expression of joy. Since each emotion has its own specific AU configuration, sign judgments correspond to the identification of a given AU arrangement allowing the recognition of the corresponding emotion. As for the message judgment approach, FEE are considered as holistic expressive displays. In other words, the expression, as a whole act and not just as a facial "surface", is a straightforward information which is recognized by an observer. Thus, emotions, but also social or cognitive information, can be inferred from this display (Scherer and Grandjean, 2008; Yik and Russell, 1999). Up to now, there is no consensus on the predominance of one process on the other (Bruce and Young, 2012) but sign and message judgment processes both result in the assignment of emotional information to the facial expression.

Different theoretical approaches aim at understanding which kind of emotional information is used to recognize the facial expression. A first approach considers that basic emotion categories are used to recognize facial expression. These basic emotions categories, as outcomes of biologically preprogrammed reactions, are clearly differentiated from each other. Ekman (1992) identifies six basic emotions: Joy, Surprise, Anger, Fear, Disgust and Sadness. However, basic emotions appear not to be independent one of another but rather related to each other (Russell, 1980). On the basis of a multidimensional scaling procedure of similarities, a two-dimensional space of FEE recognition has been proposed to give an account of the emotional information recognized in faces. According to this proposal, people are able to situate FEE on dimensional continuums which are meaningful for the interpersonal interaction such as pleasantness-unpleasantness and degree of arousal. Once again, there is no consensus on which kind of emotional information (basic emotion categories or dimensional continuums) is inferred by observers when they interpret a FEE. One must stress that this inference is only an indication of the emotion really felt by the expresser. Indeed, an emotion does not systematically involve a facial expression and a facial expression can be feigned and expressed without the corresponding emotion being felt (Reisenzein et al., 2013).

### 2.3. The manual annotation procedure

For a long time, the manual annotation procedure was basic because of, at that time, the minimalist existing technologies. Roughly, investigators – also called experimenters, facilitators or test moderators – selected pictures of FEE, and annotators – also called observers, judges or decoders – were asked to assess these stimuli. As soon as technological progress enabled it, researchers started to study dynamic stimuli, that is, videos of FEE. From then on, annotation became more specifically "the process of adding data synchronized with the stimuli", allowing the "analysis of the happenings captured" (Thomann et al., 2009).

The way of adding data with the stimuli is either discrete or continuous. A discrete annotation is a procedure in which a datum is associated to the stimulus only when the behavior is observed. It is the most used procedure because it is easy to perform by annotators and easy to analyze by investigators. A continuous annotation is a procedure in which a datum is associated to every frame or time unit of the stimulus. The continuous annotation procedure, through the gathered continuous assessments, provides a "trace" describing how the emotional states displayed on one's face rise and fall from moment to moment (Cowie et al., 2012).

Manual annotation consists of three steps: configuration, annotation as such, and data analysis (Martin et al., 2005). On the configuration step, investigators who set up the annotation experiment define relevant behaviors or events to be analyzed and how they ought to be annotated. On the annotation step, annotators assess the stimuli using the proposed configuration. Annotators can be the investigators themselves but also expert annotators (e.g., trained for facial expression recognition) or novice annotators. The third step is the analysis of the produced data. Analysis