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 can be done with the same annotation tool, or using some other data processing approaches.

Two categories of manual annotation procedures can be distinguished depending on annotators' instructions. On one hand, with free-choice procedures, annotators are asked to describe the emotion displayed by the FEE stimulus using one label or more of their own (Izard et al., 1980). Annotators' answers are then clustered by investigators to be further analyzed. The major problem of free-choice procedures is the variety of answers provided by annotators, which make them difficult, if not downright impossible, to analyze. On the second hand, forced-choice procedures request annotators to assign one label out of a set of possible labels defined by the investigators. However, an important bias of forced-choice procedures is the one of suggesting the right answer to annotators (Russell, 1993). To partially avoid prompting the correct answer to annotators, investigators can design more complex forced-choice procedures to cover the target label. For instance, rather than having annotators directly selecting one emotion label, another procedure is to chain label panels (e.g., to ask annotators first to annotate the emotional valence of the FEE and then ask them to choose an emotional label) or to propose a multiple-answer with checkboxes. Another possible bias of forced-choice procedures appears when investigators predetermine themselves annotation labels. If the right answer is too obvious the recognition rate can be artificially increased (Wagner, 1997). Even if the forced-choice procedures artificially increase annotators agreements, they have the major advantages of being time saving, of needing less observers than free-choice procedures, and of collecting enough answers for robust statistics. Each existing annotation procedure has pros and cons. Free-choice procedures are closest to the psychological process of FEE recognition but they are then biased by the inevitably flawed clustering done by investigators when analyzing annotators' answers. Conversely, forced-choice procedures allow annotators to cluster themselves the recognized emotion but recognition rates depend on the labels chosen by the investigator. To date, forced-choice procedures are mainly preferred because this procedure is easier to configure, to use and to analyze.

3. Common tools for the manual annotation of FEE

Manual annotation is a powerful experiment procedure but it is time consuming and difficult to perform. Therefore, technical improvements have been done with annotation devices (Laurans et al., 2009) and regarding stimuli complexity (i.e., visual, audio or audio-visual stimuli, Saldana, 2009; Zeng et al., 2009). Current manual annotation tools, also called CAQDAS (Computer Assisted Qualitative Data Analysis Software) have been developed to ease the manual annotation burden (Dybkjær and Bernsen, 2002; Rohlfing et al., 2006; Saldana, 2009). They are especially relevant for the study of natural FEE (i.e., dynamic and spontaneous displays of emotions) which is the direction taken by most studies on emotional facial communication. The CAQDAS can be divided into two categories. The first encloses software programs that need to define a coding grid or "coding schemes" for emotion annotation. They are called annotation software with user-defined coding schemes. The second category contains software with built-in coding schemes (called built-in coding schemes annotation software). Both types of software mainly rest on forced-choice procedures. The two next paragraphs focus on the main strengths and weaknesses of these two kinds of software (detailing their various features is out of the present paper's scope).

3.1. Annotation software with user-defined coding schemes

Software with user-defined coding schemes are all built on the same structure with a main interface in which investigators upload the dynamic FEE and define coding schemes according to their needs (Fig. 1). Each coding schemes is characterized by its temporality (e.g., instantaneous or long-lasting) and its causality (e.g., person A or B, context). Then, the annotator uses the same interface to annotate the dynamic FEE.

These tools are currently the most widely used to perform video annotation (see Table 1 for a list of the most used software). They allow both discrete and continuous annotations.

User-defined annotation software address specific needs of emotion-analysis experiments. Indeed, they have three major advantages. First, investigators can easily define annotation parameters and add coding schemes, either with the menu or directly through an "add a coding scheme" button (e.g., Advene). Second, the interface displays a temporal annotation window directly under the dynamic stimuli. This presentation allows investigators to evaluate the use of each coding scheme according to the dynamic stimuli timeline but only for one annotator. Finally, some of them have also a module for the descriptive analysis of the data.

However, despite their configurability, two issues make them difficult to use by a novice annotators in emotion recognition experiments. First, their interfaces display both the investigators' configuration interface and the annotators' buttons. Thus, in each step of the experiment, some information interferes with the task of each end-user. Second, the use of coding schemes can disturb the spontaneous annotation because they involve repeated stop and go in the video timeline to annotate them. The more coding schemes there are, the more difficult it is to annotate them simultaneously. These limitations confine their use to expert annotators and to time consuming annotations.

Moreover, in order to evaluate a consensual recognition, a high number of annotators is needed, especially for audio-visual annotations. Because annotation systems with user-defined coding schemes are difficult to use, these annotation tools often involve few expert annotators which are mainly the investigators themselves. This constitutes a main limitation for FEE studies. In order to enable novice

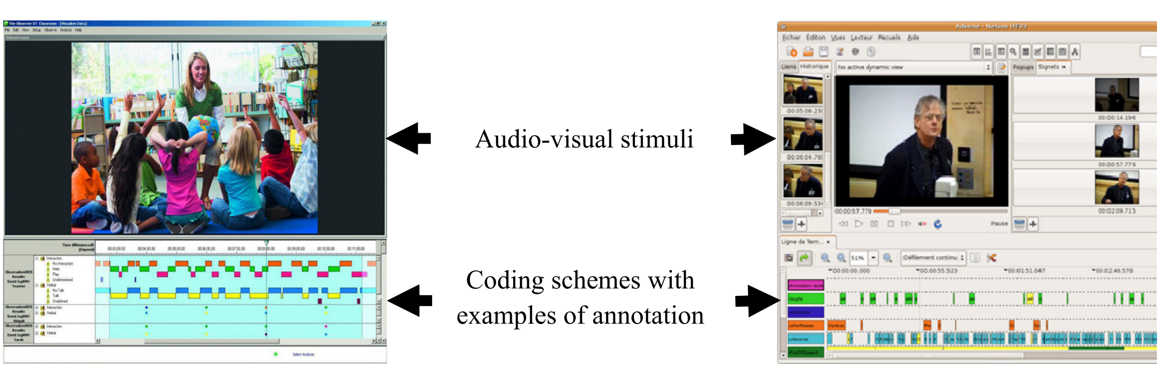


Fig. 1. Screenshots taken from the coding-scheme annotation software: Nvivo (left) and Advene (right).

Table 1
List of the available user-defined annotation software. Links were retrieved in December 2014.

Name (Ref)	Type	Link
Nvivo (Richards, 1999)	Shareware	http://www.qsrinternational.com/products_nvivo.aspx
The Observer XT (Noldus et al., 2000; Zimmerman et al., 2009)	Shareware	http://www.noldus.com/human-behavior-research/products/the-observer-xt
Videograph (Rimmele, 2002)	Shareware	http://www.dervideograph.de/enhtmStart.html
Advene (Aubert and Prié, 2005)	Freeware	http://liris.cnrs.fr/advene/
ANVIL (Kipp, 2001)	Freeware	http://www.anvil-software.org/
AmiGram (Lauer et al., 2005)	Freeware	http://sourceforge.net/projects/amigram/
ELAN (Brugman and Russel, 2004)	Freeware	https://tla.mpi.nl/tools/tla-tools/ela
Lignes de Temps (Puig et al., 2007)	Freeware	http://www.iri.centrepompidou.fr/outils/lignes-de-temps/
MacVisSTA (Rose et al., 2004)	Freeware	http://sourceforge.net/projects/macvissta/
SSI (Wagner et al., 2010)	Freeware	http://hcm-lab.de/projects/ssi/
Tatiana (Dyke et al., 2009)	Freeware	https://code.google.com/p/tatiana/
Transana Video Analysis Tool (Woods and Dempster, 2011)	Freeware	http://www.transana.org/
VCode/VData (Hagedorn et al., 2008)	Freeware	http://social.cs.uiuc.edu/projects/vcode.html
Cowlog (Hänninen and Pastell, 2009)	Freeware	http://run.cowlog.org/

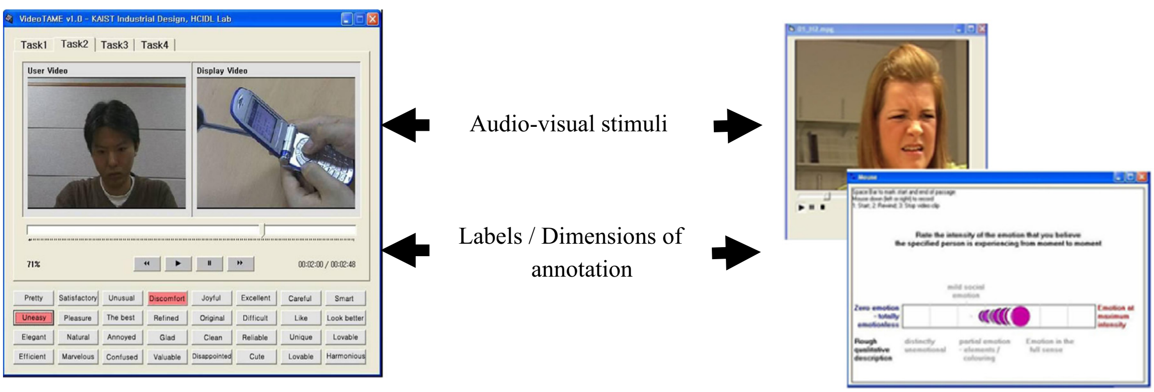


Fig. 2. Screenshots taken from the built-in coding schemes annotation software: VideoTAME (left) and Gtrace (right).

annotators to use CAQDAS, and thus increase the number of annotators, other annotation software are developed with specific built-in coding schemes.

3.2. Built-in coding schemes annotation software

Built-in coding schemes annotation tools allow novice annotators to perform annotations of data displaying spontaneous FEE with a simplified annotation interface (Fig. 2).

Some have been developed with built-in coding schemes based on the six basic emotional labels (see above), while others are based on continuous emotional dimensions (e.g., pleasure and arousal, see above). The emotional labels annotation involves defining the targeted coding schemes in a forced-choice procedure listing the basic emotions. The annotation of continuous emotional dimensions allows to rate the FEE with scales such as pleasure and arousal. Table 2 presents a non-exhaustive list of the built-in coding schemes software based either on categories or on dimensions. Categories are annotated with discrete labels whereas dimensions are annotated with continuous scales such as Gtrace (Fig. 2. right). Most of these software programs are non-commercial and thus are not still available.

As user-defined annotation software, built-in coding schemes annotation software display dynamic stimuli in the upper part of the interface and annotation interactions on the lower part. However built-in coding schemes annotation software has no interference of configuration menus, all the displayed elements being useful for annotators.

The problem of built-in coding schemes annotation software programs is their lack of configurability. Investigators cannot choose their own coding schemes because built-in software programs are not developed to be configured. They need computer scientists to modify the software according to their needs. A proposed solution to simplify the annotation process is to combine participants with system-initiated annotation such as CASAM (Computer Aided Semantic Annotation Bowers et al., 2012). However these built-in software are restricted by the labels they offer (for instance, buttons cannot be changed). Thus, they cannot be used for other types of experiments for the simple and good reason that they lack configurability.

As a conclusion, CAQDAS are relevant to perform annotation experiments. Both types of CAQDAS (user-defined coding scheme annotation software and built-in annotation software) allow standardizing the annotation procedure and computing data very precisely. However, these software have unfortunate drawbacks. User-defined coding scheme software are easily configurable by investigators but they are not easily usable by novice annotators. Conversely, built-in software are easy to use by novice annotators but they do not allow investigators to configure them according to their experimental needs. Two tools allow investigators to define in advance the labels (Actogram Kronos: Kerguelen, 2008; and On The Mark: Young, 2009) but they tolerate only very simple annotations (for review see also Frisson et al., 2010). Thus, available CAQDAS are not fully satisfying for annotation research and they should be improved to match with investigators requirements. This is why our team has developed a tool that is both configurable by investigators who are not computer specialists and that is usable by novice annotators instructed to annotate dynamic or static FEE.

4. A new tool for the manual annotation of FEE: *Oudjat* software

Oudjat is an annotation tool that is configurable according to the investigators' needs and usable by novice annotators. It is an

Table 2
List of the available built-in coding schemes annotation software programs. Some of these tools are not available now because they are in house solutions to selected projects. Links were retrieved in December 2014.

Name (Ref)	Type	Links	Coding-scheme
Gtrace software (an upgrade of FeelTrace Cowie et al., 2011, 2000)	Freeware	https://sites.google.com/site/roddycowie/work-resources	Dimensions
CARMA (Girard, 2014)	Freeware	https://carma.codeplex.com/	Dimensions
EMuJoy (Nagel et al., 2007)	Freeware	Not available	Dimensions
CVML Video Affective Annotation interface (Soleymani and Larson, 2011; Soleymani et al., 2012)	Freeware	Not available	Dimensions
EmoPlayer (Chen et al., 2008)	Freeware	Not available	Labels
Ikannotate (Bock et al., 2011)	Freeware	Not available	Labels

open source software program that can be downloaded for free on the following website: https://dynemo.upmf-grenoble.fr/. Research justifications and non-commercial use must be specified. It is compatible with Windows, Apple and Linux OS with a 64 or 32 bit version.

Two main objectives have guided the development of *Oudjat*. First, investigators often need to use easily configurable tools by their own (since all research teams do not have a computer scientist's support) to conduct discrete annotation experiments of FEE. Therefore, one objective was to provide investigators' with a flexible software program so they would be empowered to perform various basic experiments themselves without having to rely on anyone else. Second, investigators often need a simple interface for the sake of making annotators' task easier. Annotation experiments demand attention and increase annotators' cognitive load. Thus, *Oujdat* aims to provide them with an annotation interface clear and friendly to use.

4.1. Oudjat functionalities

In order to allow investigators to configure the software and annotators to easily use it, *Oudjat* dissociates the involvement of these end-users unlike the software presented above. Because investigators and annotators have different needs, *Oudjat* considers them as distinct end-users. They do not have the same expectations or skills. This explains why the system is separated into two modules (Fig. 3).

4.1.1. Module 1: Experiment configuration

The configuration module is dedicated to investigators in order to configure their experimental parameters (Table 3). Investigators first choose the annotation procedure (Step 1). This specifies the participant's task, such as a standard free- or forced-choice annotation task or a "marking sequence annotation" task which is a combination of a first general annotation task (e.g., "click when you recognize an emotion whatever it is") and a second more precise annotation task of the previously annotated sequence (e.g., "select an emotion among those proposed to describe the sequence you choose"). Investigators then define experiment parameters such as languages (Step 2), subjects (i.e., annotators) features (Step 3), experimentation features (Step 4), medias features (Step 5), medias selection (Step 6), interactions (Step 7) and instructions (Step 8). Subjects, experimentation and medias features are used before the annotation experiment to classify annotators (e.g., age and gender), experimental conditions (e.g., order of stimuli) and stimuli (e.g., emotion displayed). *Oudjat* integrates a video processor in order to select the relevant sequence of the dynamic stimuli to annotate. Navigation buttons are chosen by investigators. They decide if annotators are allowed to pause the clip, to navigate in the clip timeline or to watch it again. Investigators also define annotators' answers with buttons, checkboxes, scales, or short free labeling. A special characteristic of *Oudjat* is to chain discrete annotators' answers possibilities in order to make clear their answer (e.g., "is this emotion positive or negative?" then "select among the proposed scales to specify your answer"). The chained forced-choice annotation process can be made not only with labels but also with scales or checkboxes. There is no limit to the number of the chained forced-choice annotations. Finally, investigators specify the questions, labels and instructions intended to annotators. This first module sets all parameters that will be used by the second module to build the annotators' interface.

Another important feature of the software is the multilingualism. During the configuration phase, it is possible to define one or several languages of the interface dedicated to the annotator: In this case, each label and instruction is written in all the defined languages. In the annotation phase, investigators choose the annotator ID, the values of the defined independent variables, and the language for the concerned annotator, according to his native language. This feature allows to conduct multicultural experimentations and to compare results, considering the language as an independent variable of the experimentation.

4.1.2. Module 2: The annotation interface

The second module is dedicated to participants' annotation of the selected stimuli (pictures, videos, audio materials). It contains only relevant information to help novice and expert annotators to perform the annotation as easily as possible. The interface displays only the instructions, the annotators' tasks, and an ending message. This simplified annotation procedure allows to quickly have a large number of annotators. Then, the annotator must use the mouse to interact with the interface.

By separating the roles of investigators and annotators into two different modules, *Oudjat* solves the issue of being both easy to configurable and easy to use (Fig. 4).

With these functionalities, *Oudjat* covers a wide range of experimental possibilities. Examples of different experimental designs are presented below (Section 4.2). Experiment settings such as languages, dependent and independent variables, chosen stimuli and all experiment features are saved after the annotation.

4.1.3. Data production and analysis

During the annotation process, each annotator's interaction with the interface is recorded in an XML file, and can be exported in a.csv file. Then it can then be loaded and analyzed on most advanced statistical analysis tools, such as SPSS, Statistica, SAS or R. A data line is added to a result file each time the annotator clicks a button, fulfills a text field, or checks a level in a scale or a checkbox. The data format is the following: date and time of the user action, subject ID, values of the independent variables, language, media ID, button, checkbox or scale id, time of sequence

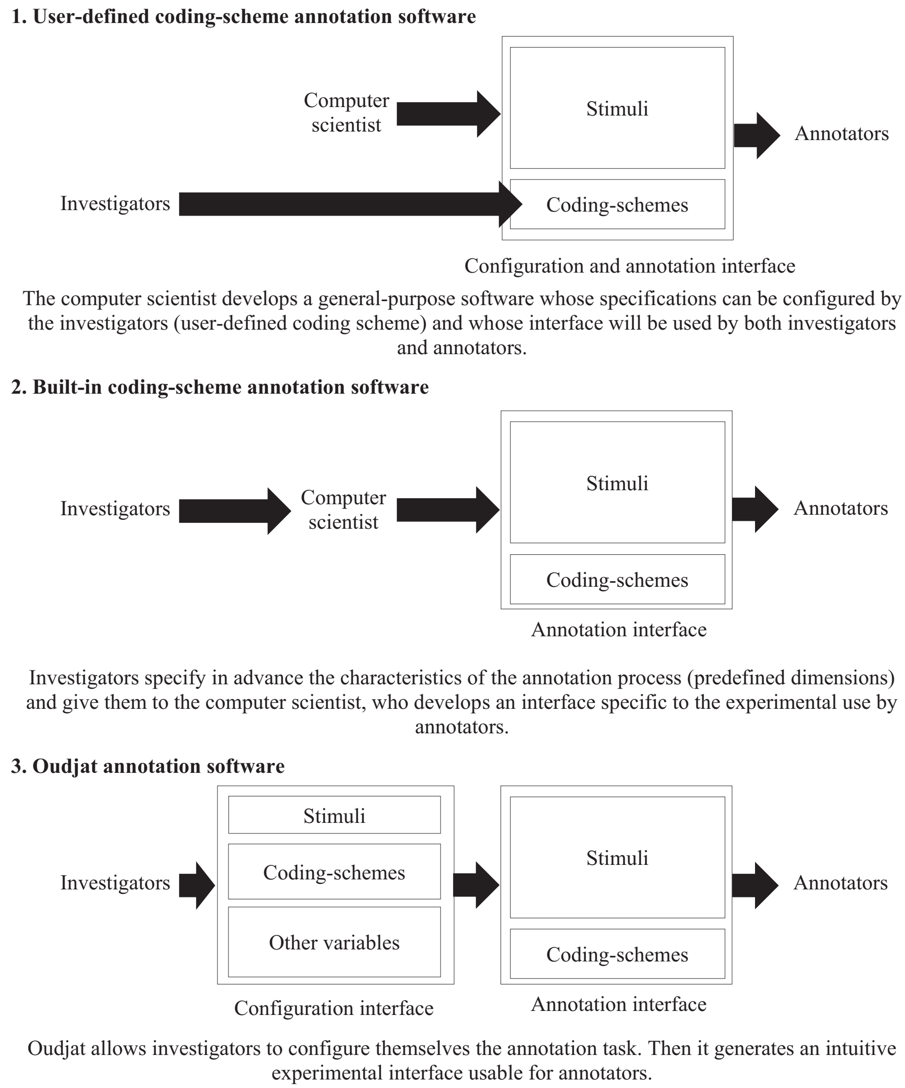


Fig. 3. A graphic comparison of software with user-defined coding scheme based annotation, built-in annotation, and *Oudjat* annotation illustrates the role of computer scientists and investigators in annotation-tool development, how investigators can configure the tool themselves to fit their own specifications and how annotators use the tools.

start, time of sequence stop (in case of sequence mark out), response time. A last column concerns an optional value field, filled with the string keyed by the annotator when choosing the "else" button, or by the level of the Likert scale measuring the annotator agreement or disagreement with the statement in 5, 6 or 7 points, if any.

4.2. Examples of emotion recognition investigations with Oudjat

As mentioned, *Oudjat*'s functionalities permit a wide range of experimental possibilities. It addresses and provides solutions for classic annotation experiments such as free-choice, forced-choice and for more complex annotation processes. Examples of experimental designs using *Oudjat* are following.

4.2.1. Discrete forced-choice annotations

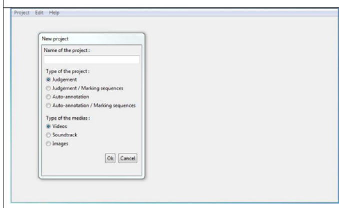
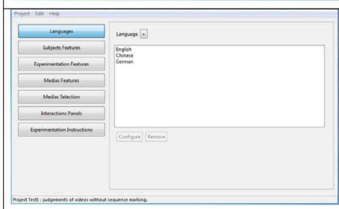
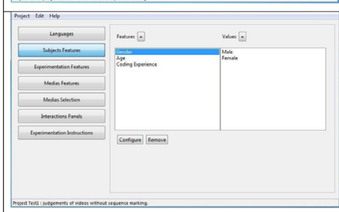
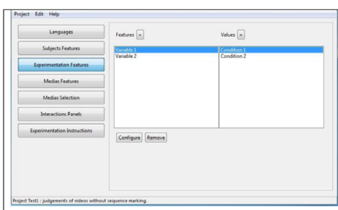
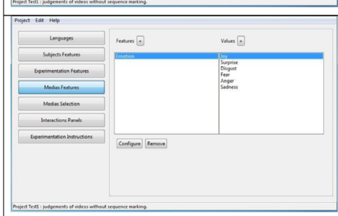
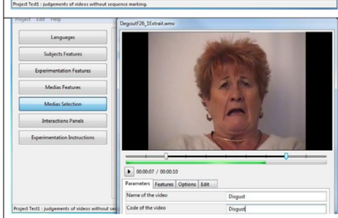
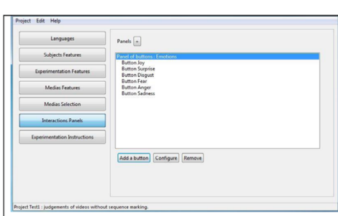
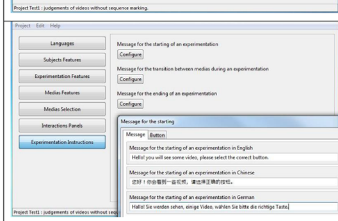
As already noted, the forced-choice procedure is the most common annotation procedure. In this design, annotators are asked to categorize stimuli by selecting an emotional label among those presented, during or directly after the stimulus display (Russell, 1993). *Oudjat* was used according to this design by Tcherkassof et al. (2007) to assess short, dynamic, and spontaneous facial expressions with few emotional labels. 67 participants were asked to assess 'on line', while watching the clip, the emotions expressed (if any) by the individuals' faces, by clicking one of eight emotional labels as soon as the target emotions were displayed (Fig. 5). Results showed that *Oudjat* permits an examination of real-time decoding activity since participants were able to attribute different emotional labels during the course of a facial expression they were looking at.

Another *Oudjat* forced-choice interface was implemented to annotate dynamic facial expression configurations depending on the parts of the face displayed (e.g., Fig. 6 only the eyes and the mouth orthe eyes, the mouth and the face, see Dubois et al., 2013 for more explainations). In this case, 242 annotators had to select a label during the emotional recognition process. The number of labels (seven) was just above the memory span in order to reduce the cognitive load and categorization errors. Results indicated that both the number and the parts of the face in complex interfaces must be considered to facilitate the emotional recognition.

4.2.2. Annotation with Likert scales or Likert items

The Likert scale is a unidimensional method. It corresponds to the sum of responses to several Likert items. A Likert item is a statement that the respondent is asked to evaluate. These items

Table 3
The 8 steps to configure an annotation experiment with *Oudjat*.

Screenshots	Description
	Step 1 Investigators choose of the annotation typology (Jugement otr Marking sequence, see section 3.2) and the type of stimuli (audio, audio-video, image).
	Step 2 Investigators define the experiment languages (here English, Chinese and German).
	Step 3 Investigators define annotators' features to sort the data (here gender, age and coding experience).
	Step 4 Investigators define experiment's variables and conditions (e.g. context of recognition, order of stimuli…).
	Step 5 Investigators define categories of the dynamic stimuli (here the emotion displayed such as joy, surprise, anger, disgust …).
	Step 6 Investigators select the simuli. A video processing can determine the target sequence to show. Investigators can also choose the presentation order (i.e. linear or random).
	Step 7 Investigators define their coding schemes, types of annotation (i.e. button, scale, checkboxes or free labelling).
	Step 8 Investigators define their instructions to start, between the stimuli and to end the experiment (here English, Chinese and German).

are usually displayed with a series of radio buttons or a horizontal bar representing a simple scale. The latter is a five (or seven or more) point scale which allows respondent to express how much they agree or disagree with the statement. An experiment on FEE contagion was conducted using *Oudjat* (Tcherkassof and Dupré, 2014). 146 annotators looked at natural FEE videos. Once the video ended, each participant filled out Likert items regarding their own emotional state induced by the video. In the present study, the annotator's emotional state was measured through dimensions. Series of radio buttons allowed them to choose the level of agreement for each statement (pleasure dimension, arousal dimension, and dominance dimension; cf. Fig. 7).

4.2.3. *Chained forced-choice annotations*

A particularity of *Oudjat* is the configuration of chain discrete forced-choice annotations in which the annotation labels leads to other annotation labels or scales. For example, in a study aimed at evaluating users feelings about an e-book reader (Sbai et al., 2010), 88 participants were filmed and asked to annotate their own emotions in a chained annotation process. To do so, participants were displayed their own face while reading an e-book and were asked to recall if they were feeling positive or negative emotions at that time. Then, a second forced-choice was presented with 12 positive or with 12 negative labels depending on the previous label choose (see Fig. 8 taken from Sbai et al., 2010). The analysis of these annotation results shows that emotional experiences are rather complex and can be composed of several emotions.

4.2.4. *Sequence marking annotation procedure*

In order to assess a FEE during its time course, a procedure consists in annotating the video stimulus without interrupting it (cf. Tcherkassof et al., 2007 and Fig. 5). Indeed, annotation procedures in which annotators stop the stimuli is problematic because, as stressed before, the dynamic is a relevant cue for FEE recognition. Dynamic information is lost when the video is stopped. 'On line' annotation avoids such problem because the participant annotates the FEE during its unfolding. This annotation can be done by asking the annotator to press a key each time he recognizes a target emotion displayed by the face (he presses the key when the emotional displays starts and when it ends, the video not being stopped). If the face displays several emotions, the annotator will press different keys, each dedicated to a given emotion. A main limitation is that only a few labels can be posted in order not to saturate annotators' cognitive load (more or less seven labels). Another main limitation is to artificially increase the recognition rates of participants because the emotion(s) to be recognized are given before, thereby guiding (even biasing) annotators' perception. Sequence marking annotation procedure circumvents such limitations by combining continuous and discrete annotations. The annotator first makes a continuous annotation by indicating when s-he perceives the beginning and the end of an emotional sequence (Fig. 8a) and, second, attributes an emotional label (Fig. 8b). In other words, during the video unfolding, the annotator first delimitates the moment when the face displays an emotion (i.e., a temporal sequence). If the video displays various emotions, the annotator delimitates several emotional periods (or temporal sequences). These marked sequences are recorded by the device. Afterwards, the marked sequences are displayed to the annotator who then makes a semantic judgment (according to a forced-choice or a free-choice procedure).

Oudjat has been configured for sequence marking annotations procedures (Fig. 9) to annotate a dynamic and spontaneous video database (Meillon et al., 2010; Tcherkassof et al., 2013). For this experiment, 171 annotators were asked to mark out the beginning

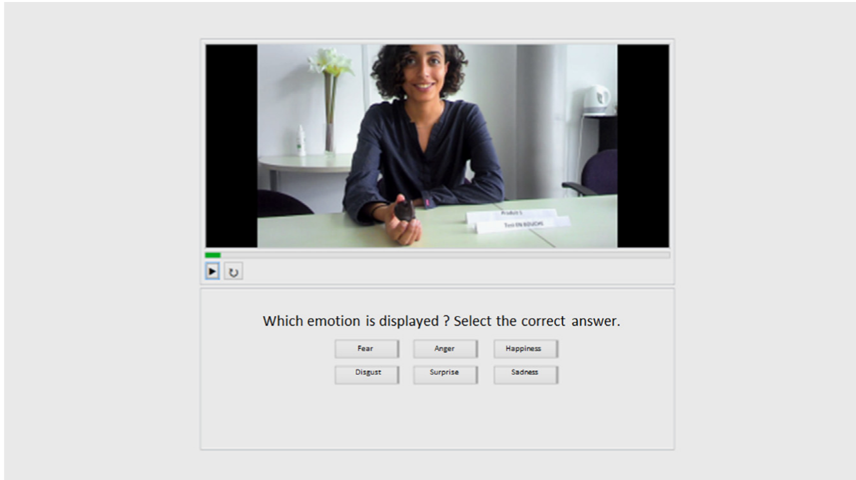


Fig. 4. An example of annotation interface with a 6-buttons panel.

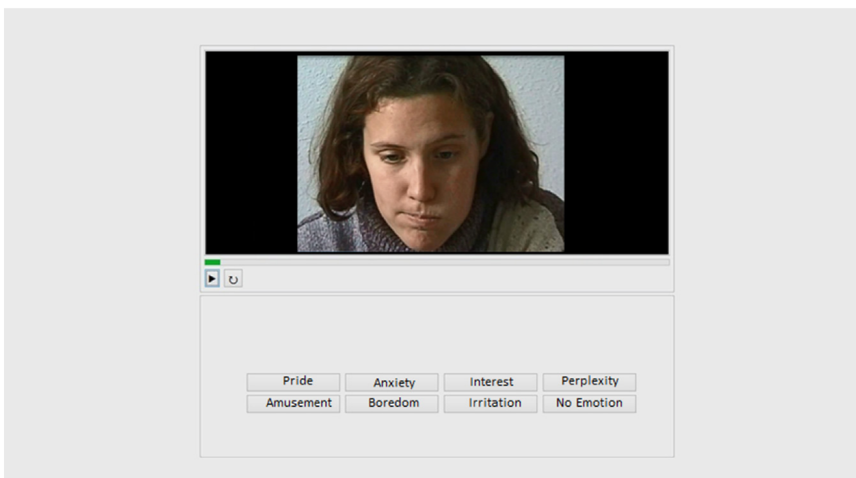


Fig. 5. The *Oudjat* configuration for emotional facial expression recognition in a standard forced choice annotation procedure (Tcherkassof et al., 2007).

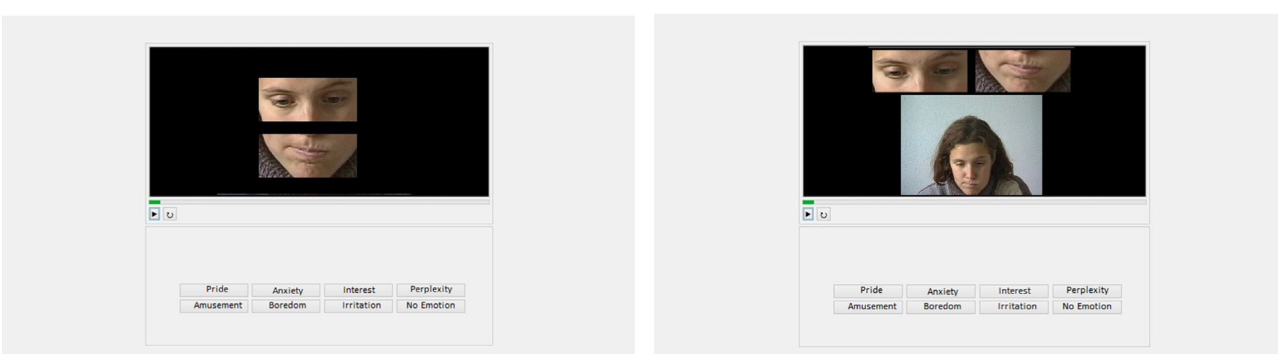


Fig. 6. Annotation interfaces of specific dynamic facial expression configuration (from Dubois et al., 2013). Examples of configurations with the eyes and the mouth only at right and with the mouth, the eyes and the whole face at left.

and end of each FEE they recognized (i.e., each emotional periods or sequences) in 33 videos. Then, they selected one (out of 13) emotional label corresponding to the expression previously marked out.

An interesting output of *Oudjat*'s sequence marking annotation is the possibility to automatically transform participants' judgments into timelines (Fig. 10). For each FEE, *Oudjat* provided a timeline showing the emotional label selected by each annotator for each marked sequence. A general timeline is then calculated by superimposing, every tenth of a second along the entire FEE recording, the timeline of all annotators. This overall timeline shows the evolution (superimposed curves) of between-annotator agreement. In other words, it shows all participants' judgments synchronized with the real-time unfolding of the FEE (and thus describes the unfolding of the facial recognition). For example Fig. 9 consists of a FEE recording of a woman during an emotional elicitation. During the first step of annotation, 100% of annotators have segmented the video from sec. 34 to sec. 54 as being

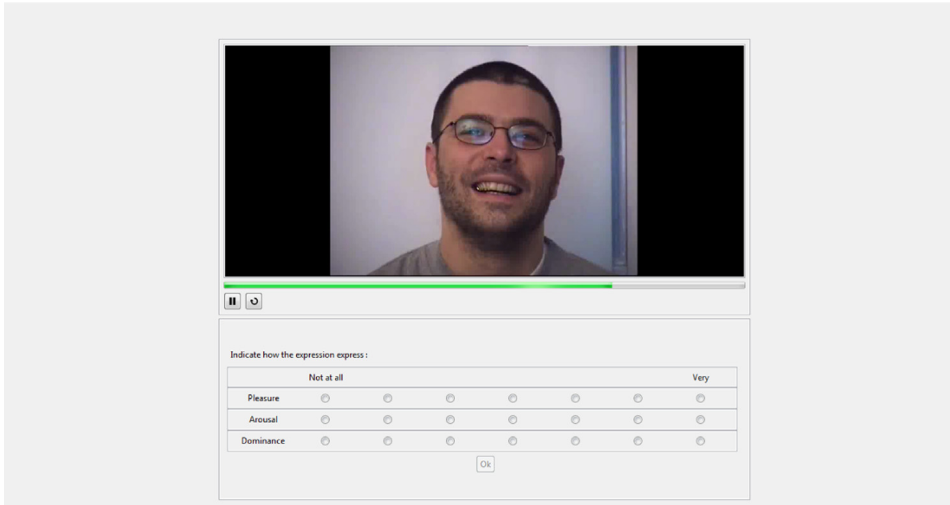


Fig. 7. Interface built by the configuration tool of *Oudjat* for Likert scales.

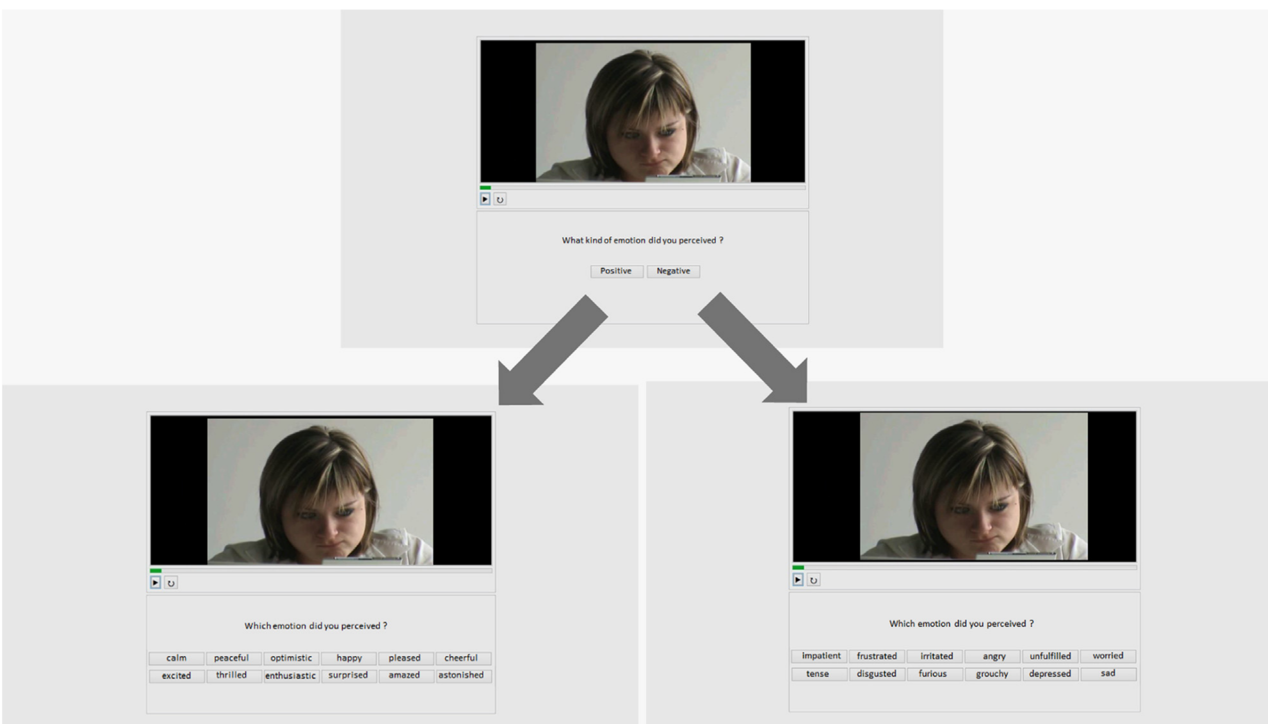


Fig. 8. The *Oudjat* configuration for chained emotion annotation during an user experience (Sbai et al., 2010). Annotators were asked to indicate first if they felt a positive or a negative emotion, and second, what emotion they felt depending on their first answer.

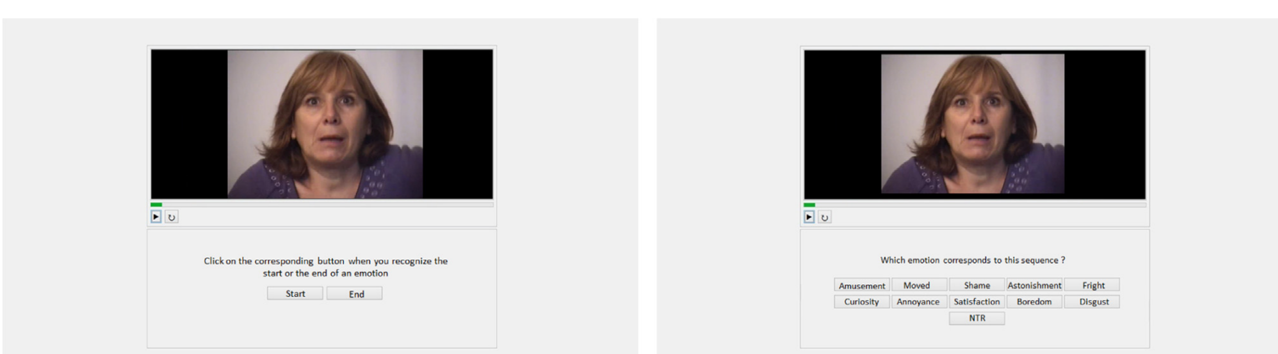


Fig. 9. The *Oudjat* configuration with the emotional sequence marking task at right and emotional label attribution task at left (Meillon et al., 2010; Tcherkassof et al., 2013).

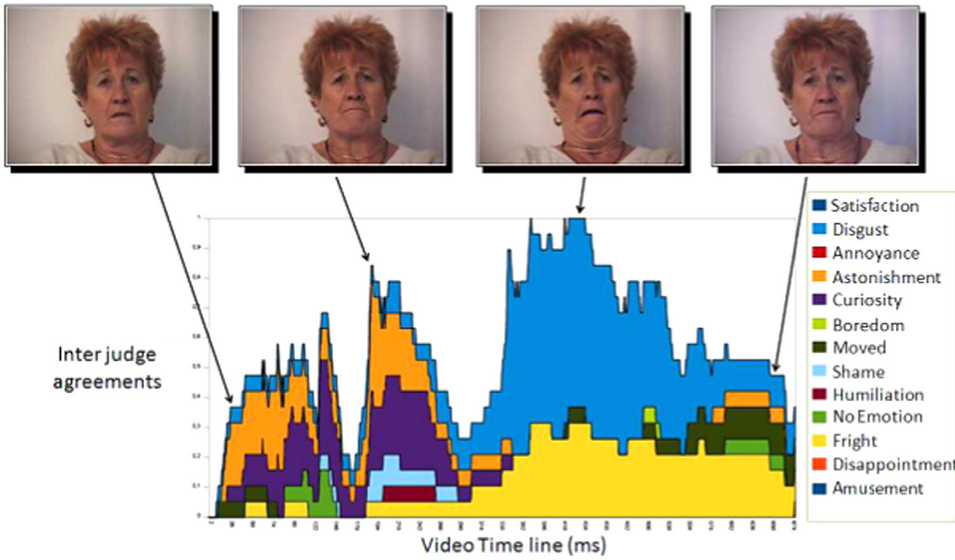


Fig. 10. Emotional expressive time-line. Frames are taken from the video of a woman who reported disgust and its corresponding underneath time-line (from Tcherkassof et al., 2013).

emotionally expressive. During the second step, 70% identified it as expressing disgust (in blue) whereas 30% have rated the face as expressing fright instead (in yellow). Thus the decoded emotions were readily identifiable. This functionality of *Oudjat* makes real “the process of adding data synchronized with the stimuli”, allowing the “analysis of the happenings captured” (Thomann et al., 2009). Identification scores can also be computed. It is done for each FEE recording by determining which frame generates the most inter-judge agreement on a given emotion label.

5. Conclusion and perspectives

The manual annotation is a relevant procedure to evaluate and to measure behaviors, emotions or perceptions (Cowie et al., 2012). Recent annotation procedures have evolved to meet the requirements of complex stimuli such as dynamic ones. However, when conducting experiments, the available tools seem to be torn between configurability and usability. Considering advantages and weaknesses of existing software programs and keeping in mind the different specifications needed for annotation experiments, *Oudjat* tool has been designed to all kinds of testing. *Oudjat* is an open-source annotation software which integrates all experimental options that might be required by various kinds of experiments. It can easily evolve with the investigator’s needs. *Oudjat* is both configurable according to complex experimental needs, and still easy to use by novice annotators. The configuration interface is available in English and in French. Yet, *Oudjat*, offers the possibility to build the annotation interface in any language, so it can be used by annotators of different native languages, in any countries. This software also offers various annotation possibilities such as free-choice annotations or forced-choice annotations with labels, scales or checkboxes. For example, it can be configured with buttons or checkboxes to reproduce classic FEE recognition tasks with the six basic emotions (Ekman, 1992). It can also be configured with other categories: appraisals (e.g., “he/she has just received a gift”, “he/she has just lost someone very close to him/her”), behavioral (e.g., “he/she wants to jump up and down”, “he/she wants to hit this person”) or physiological responses (e.g., “his/her heart is beating”, “he/she sweats”), etc. When configured with Likert scales, it is appropriate for dimensional recognition such as “pleasure” or “arousal” dimensions (Russell, 1980) or any other dimensions such as “warmth” and “competence” (Imhoff et al., 2013). It also allows investigators to chain two or more forced-choice to specify their annotation (for example, assessment of valence followed by the assessment of specific emotions). Sequence marking annotation experiments can also be conducted. In this case, annotators first delimitate a temporal sequence in the video and, second, attribute it a label. Finally, the temporal resolution of *Oudjat*’s outputs is accurate to the millisecond and can be used to produce annotator-agreement timelines by aggregating annotations (Tcherkassof et al., 2013). Despite its advantages *Oudjat* still needs to be improved. The configuration interface of the first module misses options such as keyboard shortcuts. It also lacks the possibility of continuous assessment with a joystick or other specific devices. However, since the source code is available, it can easily be reworked. Until now, *Oudjat* has been used for FEE recognition experiments. Yet, it can be extended to other kinds of emotional or non-emotional stimuli such as odor stimuli. *Oudjat*’s usefully complements each other with user-defined annotation software or built-in coding-scheme annotation software for annotation experiments. *Oudjat* can be freely downloaded from the DynEmo database website https://dynemo.upmf-grenoble.fr/.

Acknowledgments

The authors thank the University of Grenoble Alps and the Pole Grenoble Cognition for their valuable contributions and funding. We also thank the reviewers for their helpful comments.

References

Aubert, O., Prié, Y., 2005. Advene: active reading through hypervideo. In: ACM Conference on Hypertext and Hypermedia, Salzburg, Austria, pp. 235–244.

Bock, R., Siegert, I., Haase, M., Lange, J., Wendemuth, A., 2011. ikannotate: a tool for labelling, transcription, and annotation of emotionally coloured speech. In: D’Mello, S., Graesser, A., Schuller, B., Martin, J.-C. (Eds.), Affective Computing and Intelligent Interaction, pp. 25–34.

Bowers, C., Beale, R., Byrne, W., Creed, C., Pinder, C., Hendley, R., 2012. Interaction issues in computer aided semantic annotation of multimedia. In: Nordic Conference on Human–Computer Interaction. ACM, Copenhagen, Denmark, . , , pp. 534–543.

Bruce, V., Young, A.W., 2012. Recognising faces. In: Bruce, V., Young, A.W. (Eds.), Face Perception. Psychology Press, New York, NY, pp. 253–314.

Brugman, H., Russel, A., 2004. Annotating multimedia/multi-modal resources with ELAN. In: International Conference on Language Resources and Evaluation, Paris, France, pp. 2065–2068.
Chen, L., Chen, G.C., Xu, C.Z., March, J., Benford, S., 2008. EmoPlayer: a media player for video clips with affective annotations. Interact. Comput. 20, 17–28.
Cohn, J.F., Ambadar, Z., Ekman, P., 2007. Observer-based Measurement of Facial Expression with the Facial Action Coding System. The Handbook of Emotion Elicitation and Assessment, pp. 203–221.
Cowie, R., Cox, C., Martin, J.C., Batliner, A., Heylen, D., Karpouzis, K., 2011. Issues in data labelling. In: Cowie, R., Pelachaud, C., Petta, P. (Eds.), Emotion-Oriented Systems: The Humaine Handbook. Springer Verlag, London, pp. 213–241.
Cowie, R., Douglas-Cowie, E., Savvidou, S., McMahon, E., Sawey, M., Schrader, M., 2000. FEELTRACE: An Instrument for Recording Perceived Emotion in Real Time. Real Time ISCA Workshop on Speech & Emotion, Newcastle, United Kingdom.
Cowie, R., McKeown, G., Douglas-Cowie, E., 2012. Tracing emotion: an overview. Int. J. Synth. Emot. 3, 1–17.
Du, S., Tao, Y., Martinez, A.M., 2014. Compound facial expressions of emotion. Proc. Natl. Acad. Sci. 111, E1454–E1462.
Dubois, M., Dupré, D., Adam, J.-M., Tcherkassof, A., Mandran, N., Meillon, B., 2013. The influence of facial interface design on dynamic emotional recognition. J. Multimodal Interfaces, 1–9.
Dybkjær, L., Bernsen, N.O., 2002. Data, annotation schemes and coding tools for natural interactivity. In: International Conference for Spoken Language Processing, Denver, USA, pp. 217–220.
Dyke, G., Lund, K., Girardot, J.J., 2009. Tatiana: an environment to support the CSCL analysis process. In: O'Malley, C., Suthers, D., Reimann, P., Dimitracopoulou, A. (Eds.), Conference on Computer Supported Collaborative Learning. International Society of the Learning Sciences. . pp. 58–67.
Ekman, P., 1992. An argument for basic emotions. Cognit. Emot. 6, 169–200.
Ekman, P., Friesen, W.V., 1978. Facial Action Coding System: A Technique for the Measurement of Facial Movement. Consulting Psychologists Press, Palo Alto, USA.
Frisson, C., Alaçam, S., Coskun, E., Ertl, D., Kayalar, C., Lawson, L., Lingenfelser, F., Wagner, J., 2010. COMEDIANNOTATE: towards more usable multimedia content annotation by adapting the use interface. QPSR Numediart Res. Program 3, 45–55.
Girard, J.M., 2014. CARMA: software for continuous affect rating and media annotation. J. Open Res. Softw. 2, e5.
Hagedorn, J., Hailpern, J., Karahalios, K.G., 2008. VCode and VData: illustrating a new framework for supporting the video annotation workflow. In: Working Conference on Advanced Visual Interfaces. ACM, pp. 317–321.
Hänninen, L., Pastell, M., 2009. CowLog: open-source software for coding behaviors from digital video. Behav. Res. Methods 41, 472–476.
Imhoff, R., Woelki, J., Hanke, S., Dotsch, R., 2013. Warmth and competence in your face! Visual encoding of stereotype content. Front. Psychol., 4.
Izard, C.E., Huebner, R.R., Risser, D., Dougherty, L., 1980. The young infant's ability to produce discrete emotion expressions. Dev. Psychol. 16, 132–140.
Kelly, J.R., Agnew, C.R., 2011. Behavior and behavior assessment. In: Deaux, K., Snyder, M. (Eds.), The Oxford Handbook of Personality and Social Psychology. Oxford University Press, New York, NY, p. 93.
Kerguelen, A., 2008. Actogram Kronos": un outil d'aide à l'analyse de l'activité. Les techniques d'observation en sciences humaines, 142–158.
Kim, H., Somerville, L.H., Johnstone, T., Alexander, A.L., Whalen, P.J., 2003. Inverse amygdala and medial prefrontal cortex responses to surprised faces. NeuroReport 14, 2317–2322.
Kipp, M., 2001. ANVIL, a generic annotation tool for multimodal dialogue. In: European Conference on Speech Communication and Technology, Aalborg, Denmark..
Krumhuber, E.G., Kappas, A., Manstead, A.S., 2013. Effects of dynamic aspects of facial expressions: a review. Emot. Rev. 5, 41–46.
Lauer, C., Frey, J., Lang, B., Becker, T., Kleinbauer, T., Alexandersson, J., 2005. Amigram-a general-purpose tool for multimodal corpus annotation, Multimodal Interaction and Related Machine Learning Algorithms, Edinburgh, United Kingdom .
Laurans, G., Desmet, P.M.A., Hekkert, P., 2009. Assessing emotion in interaction: Some problems and a new approach. In: International Conference on Designing Pleasurable Products and Interfaces, Compiegne, France, pp. 13–16.
Lottridge, D., Chignell, M., Jovicic, A., 2011. Affective interaction understanding, evaluating, and designing for human emotion. Rev. Hum. Factors Ergon. 7, 197–217.
Martin, J.-C., Abrilian, S., Devillers, L., 2005. Annotating multimodal behaviors occurring during non basic emotions. Affect. Comput. Intell. Interact. 3784, 550–557.
Meillon, B., Tcherkassof, A., Mandran, N., Adam, J.M., Dubois, M., Dupré, D., Benoit, A.M., Guérin-Dugué, A., Caplier, A., 2010. DYNEMO: A Corpus of dynamic and spontaneous emotional facial expressions. In: Language Resources and Evaluation Conference, Malte, Malte.
Metaxas, D., Zhang, S., 2013. A review of motion analysis methods for human Nonverbal Communication Computing. Image Vis. Comput. 31, 421–433.
Nagel, F., Kopiez, R., Grewe, O., Altenmoller, E., 2007. EMuJoy: software for continuous measurement of perceived emotions in music. Behav. Res. Methods 39, 283–290.
Noldus, L.P.J.J., Trienes, R.J.H., Hendriksen, A.H.M., Jansen, H., Jansen, R.G., 2000. The Observer Video-Pro: new software for the collection, management, and presentation of time-structured data from videotapes and digital media files. Behav. Res. Methods 32, 197–206.
Pantic, M., 2009. Machine analysis of facial behaviour: naturalistic and dynamic behaviour. Philos. Trans. R. Soc. London, Ser. B: Biol. Sci. 364, 3505–3513.
Puig, V., Holland, J., Cavalié, T., Benjamin, C., Mathé, J., Haussonne, Y.-M., Liévain, S., 2007. Lignes de temps, une plateforme collaborative pour l'annotation de films et d'objets temporels. In: Conférence Francophone sur l'Interaction Homme-Machine, Paris, France.
Reisenzein, R., Studtmann, M., Horstmann, G., 2013. Coherence between emotion and facial expression: evidence from laboratory experiments. Emot. Rev. 5, 16–23.
Richards, L., 1999. Using NVivo in Qualitative Research. Sage Publications, Thousand Oaks, CA.
Rimmele, R., 2002. Videograph. Multimedia-Player zur Kodierung von Videos. IPN, Kiel.
Rohlfing, K., Loehr, D., Duncan, S., Brown, A., Franklin, A., Kimbara, I., Milde, J.-T., Parrill, F., Rose, T., Schmidt, T., 2006. Comparison of multimodal annotation tools, workshop report. Convers. Discourse Anal. 7, 99–123.
Rose, R.T., Quek, F., Shi, Y., 2004. MacVisSTA: a system for multimodal analysis. In: International Conference on Multimodal interfaces ACM, , , State College, PA, USA, pp. 259–264.
Russell, J.A., 1980. A circumplex model of affect. J. Pers. Soc. Psychol. 39, 1161–1178.
Russell, J.A., 1993. Forced-choice response format in the study of facial expression. Motiv. Emot. 17, 41–51.
Saldana, J., 2009. An introduction to codes and coding. In: Saldana, J. (Ed.), The Coding Manual for Qualitative Researchers. Sage Publications Limited, London.
Sariyanidi, E., Gunes, H., Cavallaro, A., 2014. Automatic Analysis of Facial Affect: A Survey of Registration, Representation and Recognition.
Sbai, N., Dubois, M., Kouabenan, R., 2010. Etude de l'interaction Humain-Technologie: comment mesurer les émotions de l'utilisateur?, Ergo'IA, Innovation, Interactions, Qualité de vie, Biarritz, France.
Scherer, K.R., Grandjean, D., 2008. Facial expressions allow inference of both emotions and their components. Cognit. Emot. 22, 789–801.
Soleymani, M., Larson, M., 2011. Crowdsourcing for Affective Annotation of Video: Development of a Viewer-reported Boredom Corpus. Workshop on Crowdsourcing for Search Evaluation, Geneva, Switzerland, pp. 4–8.
Soleymani, M., Pantic, M., Pun, T., 2012. Multimodal emotion recognition in response to videos. IEEE Trans. Affect. Comput. 3, 211–223.
Sullivan, G.B., Strongman, K.T., 2003. Vacillating and mixed emotions: a conceptual-discursive perspective on contemporary emotion and cognitive appraisal theories through examples of pride. J. Theory Soc. Behav. 33, 203–226.
Tcherkassof, A., 2008. Les émotions et leurs expressions [Emotions and their expressions]. Presses Universitaires de Grenoble, Grenoble.
Tcherkassof, A., Bollon, T., Dubois, M., Pansu, P., Adam, J.-M., 2007. Facial expressions of emotions: a methodological contribution to the study of spontaneous and dynamic emotional faces. Eur. J. Soc. Psychol. 37, 1325–1345.
Tcherkassof, A., Dupré, D., 2014. Spontaneous and dynamic emotional facial expressions reflect action readiness. In: World Congress on Facial Expression of Emotion, , Porto, Portugal.
Tcherkassof, A., Dupré, D., Meillon, B., Mandran, N., Dubois, M., Adam, J.-M., 2013. DynEmo: a video database of natural facial expressions of emotions. Int. J. Multimedia Appl. 5, 61–80.
Thomann, G., Rasoulifar, R., Meillon, B., Villeneuve, F., 2009. Observation, annotation and analysis of design activities: how to find an appropriate tool?. In: International Conference on Engineering Design, Stanford, USA, pp. 1–12.
Wagner, H.L., 1997. Methods for the study of facial behavior. In: Russell, J.A., Fernández-Dols, J.M. (Eds.), The Psychology of Facial Expression. Cambridge University Press, Cambridge, pp. 31–54.
Wagner, J., André, E., Kugler, M., Leberle, D., 2010. SSI/ModelUI-A Tool for the Acquisition and Annotation of Human Generated Signals. Jahrestagung fur Akustik, Berlin, Germany.
Winston, J.S., Henson, R., Fine-Goulden, M.R., Dolan, R.J., 2004. fMRI-adaptation reveals dissociable neural representations of identity and expression in face perception. J. Neurophysiol. 92, 1830–1839.
Woods, D.K., Dempster, P.G., 2011. Tales from the bleeding edge: the qualitative analysis of complex video data using Transana. Forum: Qual. Soc. Res., 12.
Yik, M.S., Russell, J.A., 1999. Interpretation of faces: a cross-cultural study of a prediction from Fridlund's theory. Cogn. Emot. 13, 93–104.
Young, D., 2009. On The Mark. URL: 〈http://onthemark〉. sourceforge. net. P 47.
Zeng, Z., Pantic, M., Roisman, G.I., Huang, T.S., 2009. A survey of affect recognition methods: audio, visual, and spontaneous expressions. IEEE Trans. Pattern Anal. Mach. Intell. 31, 39–58.
Zhao, G., Pietikainen, M., 2007. Dynamic texture recognition using local binary patterns with an application to facial expressions. IEEE Trans. Pattern Anal. Mach. Intell. 29, 915–928.
Zimmerman, P.H., Bolhuis, J.E., Willemsen, A., Meyer, E.S., Noldus, L.P.J.J., 2009. The Observer XT: a tool for the integration and synchronization of multimodal signals. Behav. Res. Methods 41, 731–735.