

# Audio Description of Emotions in Films using Eye tracking

Paula Igarada and Alejandro Maiche

**Abstract.** Dealing with the Audio description (AD) and taking into account the many works and studies carried out in the field, the aim of this paper focuses on the moment when emotions and gestures must be described —avoiding any personal interpretation by the one who describes the audiovisual material. In this sense, we will carry out an experiment evaluating the eye movement from two groups of people: one watching a film excerpt with AD and another one watching the same film excerpt without AD. The aim of this paper is to present the on-going research on the eye-tracking technology applied to the media accessibility.

## 1 INTRODUCTION

AD is a recent field within Translation Studies, though we already have access to many guidelines for doing it properly. AD is defined as a feature that makes performing arts, audiovisual products, countries' natural, historic and cultural heritage, and all kinds of socio-cultural events accessible to everyone. AD consists of “a set of techniques and abilities, whose main objective is to compensate for the lack of perception of the visual component in any audiovisual message, by providing suitable sound information which translates or explains, in such a way that the visually impaired perceive the message as a harmonious work which is as similar as possible to that which is perceived by the sighted” [1].

Taking some existing studies and guidelines as ITC [2], Benecke [3], Remael [4], Ofcom [5], Orero [6], Puigdomènech, Matamala and Orero [7], Remael and Vercauteren [8], Vercauteren [9] into consideration, it seems that there are some recommendations and ideas on what, when and how to audio describe.

When analyzing a given piece of work we have to take into account both the non-verbal visual elements and the written visual elements. When we study the verbal visual elements we have to deal with the *where* (places and spaces, dimension of the places, location of the characters, lighting, relation between the characters and the places); *when* (the time/period when everything happens); *who* (the characters in general, describing who talks, their clothing, their physical appearance, their facial expressions, body language, age, attitudes); and *what* (action, colours, music and other noises). In relation to the written visual elements, important items are the closing credits, the possible subtitles appearing on the screen, the written elements belonging to the piece of work, etc. [7]. The above studies also focus on how to audio describe, dealing with the objectivity and subjectivity during the process of AD, the elements of motion and the filmic imagery (such as symbols, associative images, etc.).

The issue that is common to all countries and guidelines developed for AD is the lack of any scientific research for its

practice; hence, there is not any agreement on how to produce good quality and meaningful AD. On one side, there is not enough research on this topic and that affects the quality of ADs. On the other side, we still find that one of the biggest challenges when doing any AD emerges when emotions and gestures must be described and the it would be desirable to avoid any personal interpretation by the one who describes the audiovisual material.

## 2 CULTURAL AND TRANSLATION STUDIES

For this study, we have considered Cultural Studies, especially in relation to Translation Studies and Studies in Psychology.

Many studies confirm that there is a series of elemental obstacles in order to have a successful intercultural perception and understanding in translation. According to Samovar and Porter [10], these factors are:

- *attitudes*, which include religion, politics, customs, food, dress, or art; the world view perspectives that may use different frames of reference; the absolute values regarding good and bad, ugly and beautiful, true or false, etc.; or the stereotypes and prejudices;
- *social organization*, as regards government, family, law, etc.
- *patterns of thought*, mainly between the Western and the Oriental worlds, but also between Anglo-Saxon and Latin ways of thinking about the family ties, moral values, dating, etc.; which cause so much confusion;
- *role expectations* obviously derived from patterns of thought;
- *conceptualization of time*;
- *handling space*;
- *language barrier*; which does not arise only with respect to correct or incorrect translation;
- and *non-verbal communication*, nonverbal stimuli in a communication setting that are generated by both the source [speaker] and his or her use of the environment and that have potential message value for the source or receiver [listener].

In this way, we can see that apart from verbal communication, physical appearance, and the use of things and the environment, among many other issues, it is possible to find other cultural signs with a communicative value. This communicative non-verbal activity within a conversation's verbal behaviour comprises a system of signs, gestures, manners, postures, ways of behaving, attitudes, etc. Some studies say that not every culture has the same way of expressing its feelings and emotions, and that not every human being uses the same gestures: Everyone experiences emotions in a particular way, depending on his/her previous experiences, learning, character, culture and situation [11, 12].

## 2.1 Non-verbal Communication

Non-verbal communication can facilitate a large amount of information in a given situation or context. Cestero [13] classifies the non-linguistic elements into the following categories: the basic systems (paralinguistic and kinesics); the cultural systems (cronemic and proxemic); and the physical and cultural signs systems.

According to Poyatos [14], the **paralinguistic** system will be determined by the non-verbal communicative activities accompanying verbal behaviour during a conversation. This system is composed of non-verbal vocal signs that work towards communication and are part of the communicative interaction. On the other hand, the **kinesic** system refers to the study of the significant body movements, including facial expressions. This system embraces the analysis of gestures, manners and postures: *Gestures* are movements with communicative value which cover facial expressions (including glances) and body movements; *manners* are the ways in which movements are expressed and they show how people do certain gestures or postures according to their culture, age, sex, emotional state, socioeconomic level, etc.; and *postures* refer to the static body positions that are able to communicate in a conscious or unconscious way. These three categories are related to each other and yet they are independent.

Poyatos also distinguishes the following cultural signs with communicative value: physical appearance, the use of artifacts and the environment. The **cronemic and proxemic** systems are part of the cultural signs with the ability to modify or add significance to the rest of the non-verbal systems or to act alone. These systems have, as a main objective, to study human beings' use and perception from space and time, respectively [13, 11].

## 2.2 Emotions and Gestures

At the beginning, we mentioned that in a conversation between at least two persons the communicative non-verbal activity comprises a system of signs, gestures, manners, postures, ways of behaving, attitudes, etc. We base our conclusion on the assumption that non-verbal communication has the ability to express a wide range of feelings, emotions and thoughts; and that there are certain emotions associated with certain gestures. In this way, we will deal with the relations between emotions and gestures in this section.

Studying works on emotions we can observe that some of the physiological and behavioural reactions derived from emotions are innate and some of them can be acquired. It appears that every emotion has its functions for different reasons, and has its consequences on the nervous system and on body expression. According to the study of Hager [15], the component that can be the core of common sense approaches to emotion is the feeling component. But this author also mentions other descriptive components of emotions such as the set of behaviours that may be performed and observed in conjunction with an emotion like the gross behaviours of the body affected by the skeletal muscles and emotion expression; the set of internal bodily changes caused by the smooth muscles and glands; the ideation, imagery, and thoughts that occur along with emotion; the circumstances that give rise to emotions; and the neural processes that underlie much of the preceding activities.

Focusing on facial expressions, Hager, between many others, shares his opinion saying that these expressions are associated

with particular human emotions and that they can provide accurate information about emotions. Hager and Ekman [16] work on how emotions and faces relate to each other, if they relate to each other at all, if the facial expressions are culturally bound or universal, etc.

Regarding the cultural relationship between emotions and facial expressions, we found works by many researchers showing the emotions on people's faces in a similar way around the world. Ekman, Sorenson and Friesen [17] and Izard [18] maintain the idea that "there are constants across cultures in the emotional meanings of certain facial expressions". On the other hand, others say that "the fact that universal connections between expressions and emotions could arise from learning which has a high probability of occurring in all cultures or from a functional role of the movements in the emotional situation" [16].

Although it is quite clear that there are universals in emotions and gestures, it is also important to bear in mind when dealing with AD that emotions and gestures have their own cultural associations and, therefore, we cannot stereotype; they are an integral part of individual and collective culture-transmitting values. If we want to fully understand the meaning of an expression, we have to be able to interpret facial expressions, gestures and body language.

In order to successfully interpreting this non-verbal communication, we focus on the possible problems we can find because of these cultural differences in body expression. According to Poyatos [14], we find out that:

- the way people make their gestures can vary between cultures;
- one gesture can be characteristic of one culture and it is possible that this gesture does not exist in the other culture; that is: it has no gesture equivalent;
- the gesture or facial expression is the same in two cultures but the meaning of this gesture is different;
- the gesture exists in two countries but it has different variants or its use is more extensive in one country than in the other one.

These are some of the problems we can find interpreting the gestures and facial expressions of people from different cultures or countries. Since we are aware that every gesture, manner or facial expression communicates an emotion, the other problem that we observe is the classification of these emotions and, even more difficult: the direct relation between one emotion and the gesture to express it.

There are studies classifying every possible human facial expression as the *Facial Action Coding System (FACS)* of Ekman and Friesen [19], that measures all visible facial movements and it estimates that there are several hundred thousand possible visibly distinguishable facial expressions, most of which are never seen on people's faces in everyday life. This system has been used to score pictures of faces that researchers have seen to express emotions and to score faces of people in emotionally arousing situations. Between their conclusions, we can find that many emotional expressions are synonyms or convey different connotations of particular emotions and that the number of expressions conveying emotional meanings is much greater than researchers thought, but it is smaller than the number of possible expressions.

Researchers continuously argue about classifications of emotions, the existence of primary and secondary emotions, that

there are studies, which say there is a series of basic emotions from which the rest of the secondary emotions originate, etc. After many readings and secondary literature, we will take for our analysis the classification of the basic emotions and all the possible “secondary” emotions that Goleman [12] facilitates. This author divides emotions into eight groups or basic families:

- *Anger*: fury, outrage, resentment, wrath, exasperation, indignation, vexation, acrimony, annoyance, irritability, hostility, and, perhaps at the extreme, pathological hatred and violence.
- *Sadness*: grief, sorrow, cheerlessness, gloom, melancholy, self-pity, loneliness, dejection, despair, and, when pathological, severe depression.
- *Fear*: anxiety, apprehension, nervousness, concern, consternation, misgiving, wariness, qualm, edginess, dread, fright, terror; as a psychopathology, phobia and panic.
- *Enjoyment*: happiness, joy, relief, contentment, bliss, delight, amusement, pride, sensual pleasure, thrill, rapture, gratification, satisfaction, euphoria, whimsy, ecstasy, and at its farthest point, mania.
- *Love*: acceptance, friendliness, trust, kindness, affinity, devotion, adoration, infatuation, and agape.
- *Surprise*: shock, astonishment, amazement, and wonder.
- *Disgust*: contempt, disdain, scorn, abhorrence, aversion, distaste, and revulsion.
- *Shame*: guilt, embarrassment, chagrin, remorse, humiliation, regret, mortification, and contrition.

The question now is **how to identify these emotions in a facial expression, in a gesture and which is the most efficient method to describe them to the target audience**. Hager and the extensive research mentioned in his work say, “The expression of a given face at a specific time is conveyed by a composite of signals from several sources of facial appearance. All these sources include the general shape, orientation (pose), and position of the head, the shapes and positions of facial features (e.g., eyes, mouth), coloration and condition of the skin, shapes of wrinkles, folds, lines, and so forth. Some of these sources are relatively fixed; others, more changeable. The most important source of change in facial expression is the set of muscular movements produced by facial muscles, which provide the most substantial changes in facial appearance over short time durations and contribute most to nonverbal communication by the face”.

Cortese and Butterfield [21] stated that there is a big problem in decoding facial-emotional expressions. Considering many studies, they carried out a code for every basic facial expression. As said by these authors, this code constitutes a system of psychological interpretation of the messages people show in their faces. This code is divided in three zones on the face: eyebrows-forehead zone, eyes-eyelids-nose zone and cheeks-mouth-jaw zone. This would be the association between facial expressions and emotions:

- *Happiness*: Eyebrows are raised up; mouth is opened and can express different kinds of smiles, the body moves, and the eyes move and have a fixed look.
- *Anger*: Staring at the offence, eyes and eyebrows have a peculiar, fixed, stiff incline, teeth may be clenched.
- *Sadness*: Lack of expression, look falls, corner of one’s mouth falls, eyebrows move, and general depression of the features.

- *Surprise*: Increase in muscle tone, interrupted breathing, dilated pupils, eyes and mouth open wide, internal and external part of eyebrows and eyelids lift up, jaw goes down, knees bend and body leans forwards.

- *Fear*: Shudder, startle, immobility for a while, possibility of falling silent or screaming, eyes and mouth open wide, corner of one’s mouth moves backward.

- *Disgust*: Eyebrows fall and come close together; cheeks, chin and upper lip go up, puckered nose, corner of one’s mouth and lower lip fall, eyes become smaller.

### 2.3 Audio description of emotions and gestures

We have to refer to the studies carried out by Salway and Graham [20] *Extracting Information about Emotions in Films* where they developed a method to extract information about characters’ emotions in films. These authors say that “the language of AD is rich in information concerning the characters and their external appearance, but information about their cognitive states, including their emotions, is not described directly. However some insight into a character’s emotional state is given by AD when the emotion is being depicted visually in the film”.

Other works associated can be the one carried out by Orero and Vercauteren [22] concerning the AD of facial expressions in animated movies and the difficulties derived from this work. They detected that facial expressions are complex especially for three reasons: the type of message that can express no emotion at all or just a mood; the fact that the face can combine more than one emotion which will make it even more difficult to come to the right conclusion; and the fact that emotions can be expressed in a more or less intense way.

Apart from these studies, we have to think about all the guidelines for AD we mentioned at the beginning of this paper: the difficulties in AD, the lack of time between dialogues, sounds, and many other elements for doing it properly. Where are the limits? At what point is it too much information when we audio describe emotions in films? When is it too obvious? How can we find a balance between these problems? How important is to describe the emotion one person has at a certain moment?

## 3 ANALYSIS

One of the biggest problems is whether AD has the ability to avoid incorporating personal interpretations or not. In this sense, a good and objective AD could be one that does not significantly alter the film-watcher’s eye movements compared to a person watching a clip without AD.

In order to obtain some objective and usable results in relation to AD of emotions and gestures we are carrying out an experiment using the eye-tracking technology with 10 participants. We will follow a similar methodology as the one developed by Masuda *et al.* [23], but instead of working with static pictures, we will use film excerpts, that is, pictures in movement; which will bring to light new and innovative results.

The procedure is as follows: these participants watch an excerpt of a film without AD and the eye movements are recorded at the same time. The main objective of our analysis is to detect the points where the majority of the group, coincides focusing their attention in the gestures and emotions.

We consider these points the places where the person resorts by the film's own dynamic: without intervention of the AD.

After that, we will take a group of people with similar characteristics and they will watch the same film excerpt but this time audio described. This excerpt contains Audio descriptions of emotions such as "she is sad" or "he feels alone". Once again, we will extract the points with a high level of coincidence.

The first hypothesis would be that there will be differences between the eye movements of the people watching the film with AD and the people watching the same film without AD: The eye movement pattern will be influenced by the AD, falling on the points that the person in charge of the AD previously selects.

In relation to this, the participants watching the film excerpt with AD will have a higher number of eye movements and saccades and the other participants will have longer eye gazes.

#### 4 CONCLUSIONS & FUTURE WORK

Once the analysis is done, we will analyze the achieved results, considering the mentioned studies, in order to obtain some guidelines, and when possible, to objectively audio describe the gestures and emotions in these films. We also take as starting point the fact that the facial expressions, gestures and all the non-verbal communication are essential in the understanding of the film.

If we observe the film excerpts chosen we find many obstacles, apart from the obvious of the AD, which can make the work more difficult dealing with non-verbal communication and emotions, such as the lack of time and the confusion with other sounds, the simultaneity of gestures and their corresponding emotions, or the fast change of gestures and, therefore, emotions, between many others.

Other problem is the description of these gestures and emotions: How should audio describers use words in this case? In an attempt of not raising once again the subjectivity or objectivity problem in AD, how should be the vocabulary, which words describe better and easily the emotions and feelings? This issue has its importance since "by only describing what we see, we are depriving the Audio description of its meaning and not creating a holistic and comprehensive Audio description" (Orero and Vercauteren, 2008).

Besides the mentioned questions, many more are appearing during the research:

- Where is the limit of the objectivity dealing with emotions in AD?
- What about the importance of the silence? Where do we have to stop and let the film/the sound go on?
- Would it be possible to translate directly the AD of a Spanish spoken language film into English, German, or any other language different from Spanish and from another culture?
- Do we all have the same way to audio describe facial expressions and emotions of people in films, and therefore, their basic emotions? The same interest?

After this first part of our study, we want to extend our study in the future with participants from other countries – including subtitles when necessary in the film excerpt – in order to prove if there are cultural differences in the perception of the emotions.

#### REFERENCES

- [1] AENOR. Norma UNE: 153020. *Audiodescripción para personas con discapacidad visual. Requisitos para la audiodescripción y elaboración de audioguías*. Madrid, Spain, p.4. (2005).
- [2] ITC. ITC Guidance on Standards on Audio Description, [http://www.ofcom.org.uk/tv/ifi/guidance/tv\\_access\\_serv/archive/audio\\_description\\_stnds](http://www.ofcom.org.uk/tv/ifi/guidance/tv_access_serv/archive/audio_description_stnds). (2000)
- [3] B. Benecke. Audio-Description. In: *Meta* 49:1, pp. 78-80. (2004).
- [4] A. Remael. Audio Description for Recorded TV, Cinema and DVD. An Experimental Stylesheet for Teaching Purposes, <http://www.hivt.be> (2005)
- [5] Ofcom. Guidelines on the provision of television access services, [http://www.ofcom.org.uk/tv/ifi/guidance/tv\\_access\\_serv/guidelines](http://www.ofcom.org.uk/tv/ifi/guidance/tv_access_serv/guidelines) (2006)
- [6] P. Orero. Algunas consideraciones sobre la audiodescripción comercial en España. In: *Sociedad, integración y televisión en España*. R. Pérez-Amat, A. Pérez-Ugena (Eds.). Laberinto, Madrid, Spain, pp. 277-292. (2006)
- [7] L. Puigdomènech, A. Matamala and P. Orero. Audio Description of Films: State of the Art and a Protocol Proposal. In: *Intermedia 2007 Lodz Conference on Audiovisual Translation and Interpreting*. Lodz, Poland (2007).
- [8] A. Remael and G. Vercauteren. Audio describing the exposition phase of films. Teaching students what to choose. In: *Trans* 11, pp. 73-94. (2007).
- [9] G. Vercauteren. Towards a European Guideline for Audio Description. In: *Media for All. Accessibility in Audiovisual Translation*. J. Díaz-Cintas, P. Orero and A. Remael (Eds.). Rodopi, Amsterdam, Holland, pp. 215-229. (2007).
- [10] L.A. Samover and R.E. Porter. *Intercultural Communication*. Wadsworth/Peter Lang, New York, U.S. (1997).
- [11] M.L. Knapp. *La comunicación no verbal. El cuerpo y el entorno*. Paidós, Barcelona, Spain. (1982).
- [12] D. Goleman. *Emotional Intelligence. Why it can matter more than IQ*. Bantam Books, New York, U.S. (1996).
- [13] A.M. Cestero. *Comunicación no verbal y enseñanza de lenguas extranjeras*. Arcos/Libros, Madrid, Spain. (1999).
- [14] F. Poyatos. *Nonverbal Communication in Translation: Theoretical and Methodological Perspectives*. John Benjamins, Amsterdam-Philadelphia. (1997).
- [15] J.C. Hager. Emotion and Facial Expression, <http://face-and-emotion.com/dataface/emotion/expression.jsp>. (2003).
- [16] J.C. Hager and P. Ekman. The Inner and Outer Meanings of Facial Expressions. In: *Social Psychophysiology: A Sourcebook*, J.T. Cacioppo and R.E. Petty (Eds.). The Guilford Press, New York, U.S. (1983).
- [17] P. Ekman, E.R. Sorenson and W.V. Friesen. Pan-cultural elements in facial emotion. In: *Science* 164, pp. 86-88. (1969).
- [18] C.E. Izard. *The face of emotion*. Appleton-Century-Crofts, New York, U.S. (1971).
- [19] P. Ekman and W.V. Friesen. *The Facial Action Coding System*. Consulting Psychologists Press, Palo Alto, U.S. (1978).
- [20] A. Salway and M. Extracting Information about Emotions in Films. In: *Proceedings of ACM Multimedia*, Berkeley, U.S., pp. 299-302. (2003).
- [21] A. Cortese and E. G. Inteligencia Emocional. [http://www.inteligencia-emocional.org/articulos/lasexpresiones\\_faciales.htm](http://www.inteligencia-emocional.org/articulos/lasexpresiones_faciales.htm) (2008).
- [22] P. Orero and G. Vercauteren. The Audio Description of Facial Expressions. In: *Audiovisual Translation: Multidisciplinary Approaches, Conference at the University of Montpellier 3*. (2008)
- [23] T. Masuda, B. Mesquita, S. Tanida, P.C. Ellsworth, J. Leu and E. Van de Veerdonk. Placing the Face in Context: Cultural Differences in the Perception of Facial Emotion. In: *Journal of Personality and Social Psychology*, 94:3, pp. 365-381. (2008).