

Describing Facial Expressions: much more than meets the eye¹

Gert Vercauteren

Artesis University College
Antwerp (Belgium)
gert.vercauteren@artesis.be

Pilar Orero

CaiaC, Universitat Autònoma de Barcelona. Facultat de Traducció i d'Interpretació
08193 Bellaterra (Barcelona). Spain
pilar.orero@uab.cat



Abstract

The article looks at one particular element that so far only got superficial attention in the guidelines: emotions and their representation on screen in the facial expressions, and the different ways in which they can be described. Facial expressions in film and other visual media constitute a very complex problem, as will be explained in the first part. Given the implicit nature of visual communication, it will not always be possible for the audio describer to determine emotions unambiguously. As the facial expressions are often shown in (short) close-ups, time to describe them is very limited and concise description is a must. In the second part of the article we look at how this can be achieved, and we briefly look into a longstanding bone of contention in the field: the level of interpretation needed and allowed in this type of description. In order to see what specific problems beginning describers experience with this regard and with a view to improving (academic) training in this field, we also conducted a small-scale experiment at UAB. In the conclusions we suggest some further avenues of research and briefly look at what all this could mean for training in audio description.

Keywords: Audio description; facial expressions; emotions; cognition; accessibility.

Resumen

El artículo analiza las emociones y su representación como expresiones faciales, y las diferentes formas en que pueden ser audiodescritas. Las expresiones faciales en el cine y otros medios visuales constituyen un problema muy complejo, como se explica en la primera parte del artículo. Dado el carácter implícito de la comunicación visual, puede que no siempre sea posible para

1. This research is supported by the grant from the Spanish Ministry of Finance and Competivity no. FFI2012-39056-C02-01 «Subtitling for the deaf and hard of hearing and audio description: new formats», and also by the Catalan Government funds 2009SGR700.

el audiodescriptor determinar sin ambigüedades las emociones que se deben describir. Las expresiones faciales que se observan a menudo en los primeros planos también son difíciles de describir y en ocasiones el espacio es muy limitado y debe proporcionarse una breve descripción. En la segunda parte del artículo veremos cómo se puede lograr, y analizar el nivel de interpretación necesario y permitido en la audiodescripción. A partir de un experimento de descripción, se estudiaron las estrategias utilizadas más frecuentemente. El artículo concluye sugiriendo nuevos temas de investigación y su aplicación a la formación en la audiodescripción.

Palabras clave: audiodescripción; expresión facial; emociones; cognición.

Summary

- | | |
|--|---|
| 1. Introduction | 6. The reception and re-production of man-made emotions |
| 2. Looking at facial expressions | Conclusion |
| 3. Basic emotions and their description | Bibliography |
| 4. The nature of facial expressions | |
| 5. The intersemiotic nature of the description process | |

1. Introduction

In his article on the discourse of pictures, Prince (1993) writes that part of the power and the appeal of movies lies in the ability of the medium to faithfully represent the subtleties of facial expressions. The accurate representation of a character's state of mind or emotions undoubtedly elicits the involvement of the audience, and contributes to the «J. viewer's understanding in cognitive and affective terms of the meaning of scenes depicted on screen» (1993: 24), but there is another explanation for the 'popularity' of the many close-ups of faces and other expressive devices in films (cfr. Fresno (forthcoming), Fryer (2010), Mascarenhas (2009)).

Most Hollywood movies are based on stories that unfold as a continuous chain of cause and effect, and the actions and events that drive the development of this chain have to be motivated. As pointed out by Bordwell (1985a: 13) «Coincidence is especially dangerous in this context, and Hollywood rule-books insist upon confining coincidence to the initial situation». In other words, a story can be triggered by a coincidental event —e.g. two characters can meet by accident— but the resulting effects and actions will usually all be motivated. Moreover, although causes can be of an impersonal type, such as the thunderstorm causing the plane crash bringing Chuck Noland (Tom Hanks) to a deserted island in *Castaway* (Zemeckis 2000), much more often the causality in Hollywood features is character- and/or emotion-driven.² Characters almost always strive to overcome obstacles and achieve goals, meaning that we are dealing with a psy-

2. As a matter of fact, the rest of the story of *Castaway* mainly focuses on Chuck's attempts to get off the island, hence displaying a personally motivated development.

chological causality; even though the intensity and stability of the characters' emotions vary with film genre, «personal or psychological character-driven causality is the armature of the classical story» (ibid.).

When the narrative of a movie makes characters the main driving force behind the story, it has to equip them with a relevant and coherent body of character traits and qualities. Here lies an essential difference between the verbal—in particular written—representation of a story (e.g. a novel) and its visual counterpart (e.g. a play or a movie) that is particularly relevant for audio describers.

Whereas the author of a novel has unlimited time and place to work out the psychology of his characters, creating very complex and multi-faceted entities through elaborate descriptions, screenplay writers, film directors and actors are working in a much more constrained environment. Time is usually limited to between 90 and 120 minutes, and if the actions shown on screen are to be motivated personally and/or psychologically, the audience has to get a general idea of the characters' state of mind as quickly as possible. These limitations could explain why films often rely—probably on a more explicit level than written fiction—on rather stereotypical models for characterization.³ As Porter-Abbot (2002: 109) put it, from our cultural and social background, «we draw upon pre-existing *types* [...] out of which, guided by the narrative, we mentally synthesize, if not the character, something that stands for the character.» In other words, the basis for most movie characters is one of these known types (the die-hard detective, the naïve blonde girl,...) to which a few 'unique' traits are added. In many instances, this limited characterization requires a higher degree of inferencing from the audience (Orero forthcoming), and thus also from the audio describer, leading to a double problem: as some of the character-related information is implicit and has to be inferred, there will always be room for ambiguity, and determining (and describing) it will always include a personal and uncertain component. Moreover, as in audio described movies different semiotic channels are competing for the target audience's attention, increasing their inference load, the audio describer will have to decide whether he wants his audience to infer characters' traits and emotions in 'the same way' he did (i.e. more inferencing) or lower the load by making his own inferences explicit in the description. Either way, descriptions of anything related to psychological characterization will have to be as complete and accurate as possible.⁴ But in films, as

3. In this respect, it is also interesting to keep in mind that not all audiovisual productions share the same filmic language, or as Monaco (1977: 542) put it: «Television is not only better equipped than most other media to deal with subtle development of character; it is also conversely poor equipped to succeed with other basic dramatic elements. Because it is much less intense than cinema (it gives us less visual and aural information), action and spectacle come off more poorly than in the movie or theatre. And because it is certainly less intimate than live theatre, it can't deal as well with the high drama of ideas and emotions.»
4. The popular belief of «a picture is worth a thousand words» can provide the audience with an endless interpretative possibility. Hence, descriptions should be as precise as possible to avoid this production of a myriad of images. Politically correct language, for example, may lead to euphemisms, and a far fetched and long winded route to assign a signified to the signifier. In Pragmatic terms this lack of definition or overgeneralisation will break the principle of cooperation in most of Grice (1989) maxims, hampering the communication.

already mentioned, time is a key restriction and the describer will have to make a selection of what he describes and will have to decide on how to describe it in order to provide the most information in the shortest possible time.

Films have a vast array of possibilities to express characters' feelings and emotions. They can verbalise them in the form of a dialogue, or represent them in an 'internal monologue' expressing what a character is thinking at a particular moment. Verbalised renderings of feelings or emotions can be important for the development of the visual narrative and plot, since it will be essential information anchoring the narrative, but it is not the main concern of our article, because this verbal content is already accessible to the audience and does not need to be described.

On the other hand feelings and emotions can be expressed in non-verbal ways too,⁵ for example through gestures, body movements, facial expressions, etc.⁶ In those cases it is the describer's task to produce an accurate verbal translation of this visual representation of feelings that are not explicitly expressed by the dialogue. Although in many instances in movies, facial expressions are intrinsically linked to and shown in combination with other body movements. In this study it was decided to focus exclusively on the face. This choice was triggered by the fact that faces offer the most universal way of expressing feelings and emotions, moreover, it has been demonstrated by Mital *et al* (forthcoming), and Orero & Vilaró (forthcoming) that it is in the face where the highest semantic complexity of any scene resides, and that it is the face on which audiences focus first and longest.

2. Looking at facial expressions

Stemming from a need to classify and understand them, facial expressions have been studied for many centuries and from almost every perspective, ranging from Theatre and Dance to Psychology, Ethnology, Anthropology, or Medicine such as the taxonomy in Figure 1 from *Mécanisme de la Physionomie Humaine* (1862).

One of the most influential studies in recent years is that in psychology from Ekman & Friesen (1971), who developed the so-called Facial Action Coding System (FACS) to classify every possible human expression. Their system has been refined to make it applicable to all cultures allowing for intercultural comparison. This international approach has made it possible to use the FACS in practical applications in many areas, but the one which has hit the press — and

5. Film makers can use specific techniques, camera movements, light and contrast to highlight or intensify emotions (Douglas & Harden 1996, Lehman & Lhur 1999). In addition, diegetic music and sound effects can convey emotions as well (Remael (forthcoming); Igareda (forthcoming)).
6. We are aware of the distinction between the natural bodily, real-life communication of feelings and emotions and the posed behaviour in movies, but we do not think that the expression as it is shown on screen is any different from the natural one. On the one hand, actors learn how to act with their bodies and faces and there are countless handbooks telling them how to reproduce a certain feeling or emotion as naturally as possible. On the other hand, scholars studying emotions and facial expression very often also use pre-produced material in their laboratory experiments, and afterwards describe and present their results as useable in real life.



Figure 1. Duchenne’s *Mécanisme de la Physionomie Humaine*.

more poignantly since 9/11 — is that of understanding people’s emotions when being interviewed by police or immigration officials in US. The wealth of information that can be obtained from facial expressions has triggered a vast array of research in Visual Computing, see Figure 2, where the objective is to create a computerised system to detect and track human expressions. One of the models developed is the Candide grid, in Fig 2. This grid was developed to code human faces according to certain parameters. Looking at the grid below, we can appreciate the triangles which allow for face animation by computer, creating emotions.



Figure 2. The CANDIDE geometric grid overlaid on the corresponding face.

Tracking facial expressions is considered an important task in IT for the development of human-centred interfaces and video analysis (Tian, Kanade & Cohn 2011). We live in a society where human-machine interaction is ubiquitous and developing a good system is essential to recognize and, subsequently, clone facial expressions. It comes to no surprise that tracking and labelling and tagging human expressions in order to describe emotions will also be of interest in AD.

3. Basic emotions and their description

As Ekman and Friesen (2003) point out, it is very difficult to find a specific body movement that always signals fear or anger, but there are distinct facial patterns for all emotions which are common to all cultures. Moreover — and this may also help describers choose the correct formulations for the various emotional states expressed by the characters — they go on to explain that the body does not really show our emotions, but rather how we are coping with them, and that the human voice reflects emotions as well, which also has a direct implication for the right interpretation and description. In other words, when describing certain facial expressions, the describer will have to look for words translating the person's actual feelings, whereas when describing body behaviour, he will have to translate the person's reactions to these feelings, since the body reflects our dealing with the emotion.

Traditionally, AD guidelines and personal notes from professionals offering training state that facial expressions have to be described in an objective way, i.e. describing what can be seen on screen and avoiding any kind of personal interpretation, in order to allow the blind audience to recreate the characters' emotions for themselves. To some extent this is possible, since different studies have proved that some emotions are universal, and are to a very large extent recognised correctly by non-trained observers (Ekman, Friesen and Ellsworth 1972). Since it is possible to recognise these emotions, it should also be possible — with proper training — to describe them objectively. Moreover emotions and facial expressions are not produced in isolation, and as movies are usually redundant in the cues they offer, the other semiotic channels provide a lot of additional information allowing the audience and the describer to make adequate inferences.

On the other hand however, the describer should be careful not to take this objectivity rule too far. Beginning describers might find it difficult to decide



Figure 3. Surprise captured by Ekman (2003).



Figure 4. Anger by Ekman (2003).

where to draw the line, and if they adhere to the recommendations too strictly, the audio description for the face seen in Figure 3 could end up as: «her upper eyelids and brows rise; her jaw drops open», or Figure 4 as: «her eyelids tighten as her brows draw together and she presses her lips together.»

Although these descriptions accurately mention the many muscles used to modify the facial features (describe what you see), it would be very hard for the audience to derive the ‘correct’ emotion from this description. Moreover, because we, as sighted viewers, understand the emotional meaning of these changes in the face (cf. above), it would make sense to describe Figure 3 as «the lady is surprised» or Figure 4 as «the lady is angry». This seems to offer various advantages: first of all, these descriptions are shorter, so more time, a much sought after commodity in AD, will be available for adding other information which may also complete the picture presented in the image. And because the describer already derived the underlying emotional meaning of the facial expression shown on screen and translated it in his description, the inference load on the part of the blind audience will be lowered.⁷

4. The nature of facial expressions

Notwithstanding the fact that basic emotions are generally recognized correctly, even by untrained observers, some elements might complicate the describer’s task. First of all, the face is a very complex source of information that sends different kinds of messages.⁸ Facial expressions can be neutral or show one of the different so-called universal emotions.⁹ Facial expressions may also present other information, for example expressing a mood or an attitude.¹⁰ Hence the describer will not just have to look for well-known emotions but has to keep in mind that many other signals may be shown by the face. This last fact may partially explain the issue raised above regarding the objectivity or subjectivity of facial expressions: whereas fear or anger will generally be recognised as such by most observers, it will be much harder to find agreement on, for instance pity or jealousy,

7. Although, for as far as we know, no research has been done on the cognitive load audio descriptions put on the blind audience, it is clear that they are presented with a myriad of stimuli to be processed, e.g. dialogues, sound effects, description of images in the case of audio described films of TV programmes. As watching these films or programmes is supposed to be an enjoyable activity, it would make sense not to unnecessarily increase the inference load on them.
8. Attention should be paid to the fact that not everybody sees the same, from a physiological stand (Fresno forthcoming, Marchant *et al.*, 2009).
9. Most research seems to agree that there are at least six universal emotions: sadness, anger, joy, fear, surprise and disgust/contempt (Faigin, 2003). Ekman and Friesen (2003) recognize that there might be other categories, such as interest, shame or excitement, but add that the universality of these latter emotions has not yet been sufficiently studied.
10. Both mood and attitude are closely related to emotions, but differ from the latter in that they last longer. When somebody is angry for a few minutes, we would call that an emotion, when the anger lasts a whole day, we would say that person is in an angry mood (see Ekman and Friesen, 2003 pp. 10-14 for a more complete account of the difference between these three concepts and the other signals that can be reflected by the face).

expressions Faigin (2003) calls subjective and circumstantial. Faigin himself is a painter and in his work he describes techniques for drawing or painting still pictures. So when he says «take a picture — of the face only — that you are convinced shows any of the above feelings,¹¹ show it to twenty people [...] and chances are you will get at least fifteen different responses», he is right in calling them subjective. However, he immediately goes on to say that these feelings are at the same time circumstantial and that they need a context to pinpoint their exact meaning. And given the motivated, character-driven development of movies, there will almost always be a context available to decide what feeling is being shown.

5. The intersemiotic nature of the description process

In addition to these inherent elements that can complicate recognition and interpretation of facial expressions, there are other external factors that render the describer's task more difficult. Those are on the one hand related to the nature of the filmic medium and on the other to the intersemiotic process of the description itself.

As pointed out in the first part of this article, movies are multimodal products that create meaning in a complex variety of ways, some of which are visual, others which are verbal or aural. In his script, the describer has to focus on the meaningful elements within the narrative content that are not readily accessible to the blind and partially sighted viewer, i.e. all the purely visual elements and those verbal (e.g. subtitled foreign language dialogue) and aural elements (e.g. sounds) whose meaning will not be clear without an additional explanation. Moreover, he has to fit in his descriptions in the intervals in which the production is 'silent', i.e. at moments when no other meaningful information is presented aurally. This means he will have to interweave his description with the dialogues and sound effects, while interacting with them, and will very often not have much time to describe what he sees, so that he has to summarize the full meaning of the expression shown on screen, which could be one of the positive results of describing emotions and not facial expressions, as already suggested in section 3 above.

This rule to describe only when there is no dialogue or other functional sound effect, complicates the description in another way as well. Very often, dialogue sequences are filmed in such a way that the characters' faces and the emotions shown on their faces are highlighted. Not infrequently the expressions shown at those times are as important and telling as the actual dialogues — especially when there is a contradiction between them as in the case of irony. In those instances the describer has no choice but to give priority to the dialogue and possibly insert a description later. It goes without saying that it is especially in these cases that information about a character's psychology is lost.

11. In addition to the already mentioned pity or jealousy, he names, among other feelings, reverence, greed, vanity, suspicion and disappointment.

6. The reception and re-production of man-made emotions

The description of emotions through facial expressions is an important skill describers should master, and given its far-reaching complexity, it should be present in any audio description training.

In preparation of the development of possible exercises for understanding and describing facial expressions, a simple test was devised for the students from the MA in Audiovisual Translation at Universitat Autònoma de Barcelona in May 2008. The subjects were a class of ten students who had followed a ten-week course in Audio Description. The material for the test was Pixar's *Knick-Knack* (Lasseter 1989) —a computer animated film. This choice was made because:

- a) it was computer created with man-made emotions —as opposed to spontaneously generated emotions. That is we didn't want a film-star producing facial expressions which may have been difficult to understand. We wanted some stereotyped facial expressions which had been already chosen by the Pixar producing team to represent certain emotions, the reception of which was supposed to be unequivocal for viewers across the world.
- b) there was no dialogue, so no verbal clues, distraction or influences.
- c) there was no music to intensify or hint at an emotion, the background music is the same throughout.
- d) the duration of the short was three minutes
- e) there was a simple but clear linear plot and story
- f) even though it was an animated film, the characters represent humans, unlike in other animated films such as *Cars* (Lasseter 2006), or *Wall-e* (Stanton 2008) where objects have human traits.

The test took place after the course had finished, and students were aware of audio description conventions and restrictions. The instruction they were given before starting the exercise, was to describe the film, not taking into consideration the time —hence allowing them certain freedom to produce full descriptions.



Figure 6. The crowd of characters from *Knick-Knack*.

The film tells the story of some souvenirs on a shelf. From Miami there is a little swimming pool with an insinuating blonde sunbathing in a small bikini. From Alaska there is a snow-ball with an igloo where Frosty the snowman lives. The story narrates the many attempts — and failures — by Frosty trying to get out of the ball to join the blonde from Miami and the rest of the souvenirs.

In Figure 7 we can first see Frosty who is about to use the pneumatic drill, and then the scene following his botched escape attempt. In the first scene, his face sports a wide smile made out of coals. Students described the actual facial expressions of the snow-man, e. g. «he has a wide smile» rather than his emotion «he is happy with his brilliant idea». Every time the snow-man fails to escape from his snow globe his mouth drops, and again students described the facial expression rather than the feeling of frustration.

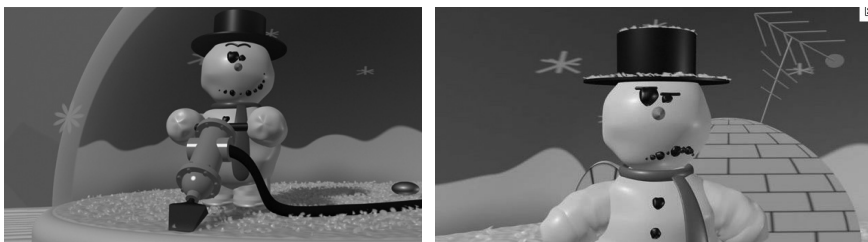


Figure 7. Happy and angry Frosty attempting to escape the snow globe.

After the experiment the students were interviewed. They described the facial expression because they had been instructed to be «objective» as most audio description guidelines suggest. Further questioning revealed students understood the feeling — if not from the expression alone, then with the help of the immediate context — but had difficulties verbalising what they saw.

This is one of the specific difficulties of translating the meaning of visual images into the verbal mode, an issue also raised by Kress (2003). Whenever you describe what you see, there is the need to verbalise, to lexicalise the picture, a problem that is intensified in the case of facial expressions as explained by Prince (1993) referring to a study by Birdwhistell (1970). Birdwhistell studied body motion cues (including facial expressions) as a form of communication in everyday life and discovered that language is not the most important channel of communication, but is only one part of a multi-modal and multi-sensory process. The specific ways in which the different modes are used is culturally, or in the case of some emotions even pan-culturally, determined and is not the object of a conscious learning process. As Holsanova (2008: 20) points out «The way we create meaning from our experience and describe to others can be understood in connection to our general communicative ability». This might explain why students recognize the expressions but do not have the words for describing them readily available (and in this respect, the case of facial expressions is not the only exam-

ple). The audio describer should bridge the gap between the iconic and the symbolic meanings of an image. That is he should go beyond the principle of mere photographic enumeration by description towards the level of metaphor and metonymy. The viewer draws conclusions from the stimuli, often inductively, on the basis of «premises furnished by the data, by the internalized rules, or by prior knowledge» (Bordwell 1985: 31), that is cognitive data. All the information we gather from top down mental processes, that is beyond physiological perception, is organised in clusters of knowledge: schemata. And it is these schemata that guide each hypothesis confirmation/deletion process needed to reach an input. These schemata may be of many kinds such as

- Prototypes, for example blonde women
- Templates, for example a linear telling of a story
- Procedural patterns, such as riding a bicycle, which are behavioural skills

Schemata play a crucial role in story comprehension and since many of the inferences and hypotheses are tested through prior knowledge, expectations, etc. it means that cognition is an acquired skill. And as with any activity that can be acquired, we can train students to contextualise; since schemata are stored and can be used in future cognition. As a viewer constructs a wider repertoire of schemata, the process of testing and challenging other possible meanings becomes richer and fuller. This has a direct implication for the viewer/audio describer because it means people can be trained to see and also to understand the meaning of what they see, even though the results from the training are perceptual and conceptual abilities which are subtle and nuanced hence difficult to evaluate. In a later phase then, training is also needed to teach students how to reformulate this visual understanding as accurately as possible.

It would go beyond the scope of this article to elaborate on all the implications of this observation for AD training programmes, but it suggests that the acquisition and constant development of a varied and extensive active, visual, and abstract vocabulary — for both recognition and description — should be a central part of these programmes.

Conclusion

As research into audio description starts to develop, it becomes ever more clear that existing guidelines, valuable and indispensable as they may be, should not always be taken at face value. They leave some basic questions unanswered and, as we tried to show in this article, taking them too literally could result in descriptions that are far from ideal. The audio description of facial expressions is much more than a choice between ‘objective’ and ‘subjective’ wording. These expressions and the emotions they portray, represent an intricate complexity that cannot be overlooked. Both internal factors such as the inherent ambiguity of visual images and the myriad of other signals conveyed by the face, and external factors such as the way they are shown on screen (short close-ups) and the difficulty of the intersemiotic transla-

tion process, make them one of the elements that are hardest to describe adequately. Moreover, as the short experiment we conducted showed, accurate verbalisation of visual images seems to be a problem that adds to this complexity.

This article is only a very first step and much more research is needed in this field. For instance reception studies are needed to find out what kind of descriptions audiences prefer (and in what context, as in some instances more objective descriptions may be OK while in other they would be too hard to decipher correctly); another, probably much more difficult, avenue of research, could consist in testing the inference load the different types of descriptions present.

As far as training is concerned, it is clear that there is room for improvement here as well. First of all cognitive skills have to be developed so beginning describers learn how to recognize and understand facial expressions correctly (including an awareness raising component showing that not all expressions are unambiguous and there is much more to facial expressions than meets the eye); further there seems to be an urgent need to develop their verbalisation skills, especially because this is one point where existing guidelines are very clear and right: audio description language should be vivid, varied and eloquent, and that certainly goes for the crucial aspect that is the description of facial expressions and emotions.

Bibliography

- BALLESTER, A. (2007). «Directores en la sombra: personajes y su caracterización en el guión audiodescrito de *Todo sobre mi madre* (1999)». In: Jiménez, C. (ed.). *Traducción y accesibilidad. Subtitulación para sordos y audiodescripción para ciegos: nuevas modalidades de Traducción Audiovisual*. Frankfurt: Peter Lang, p. 133-151.
- BIRDWHISTELL, R. (1970). *Kinetics and Context*. Philadelphia: University of Pennsylvania Press.
- BORDWELL, D., et al. (1985a). *The Classical Hollywood Cinema. Film Style & Mode of Production to 1960*. Oxon: Routledge.
- BORDWELL, D. (1985b). *Narration in the Fiction Film*. Wisconsin: The University of Wisconsin Press.
- DORNAIKA, F.; BOGDAN, R. (2007). Efficient Facial Expression Recognition for Human Robot Interaction. *LNCS*, p. 700-708. <http://www.springerlink.com/content/g055tg-v261nj1472/> [consulted 02/04/2010]
- DOUGLASS, J. S.; HARDEN, Glenn P. (1996). *The Art of Technique: An Aesthetic Approach to Film and Video Production*. Needham Heights MA: Allyn & Bacon.
- DUCHENNE, G.-B. (1862). *Mécanisme de la physionomie humaine, ou analyse électro-physiologique de ses différents modes de l'expression*. Paris: Archives Générales de Médecine, P. Asselin; vol. 1: p. 29-47, 152-174.
- EKMAN, P.; FRIESEN, W. V. (1971). «Constants Across Cultures in the Face and Emotion». *Journal of Personality and Social Psychology*, 17: p. 124-129.
- EKMAN, P.; FRIESEN, W. V.; ELLSWORTH, P. (1972). *Emotion in the human face: Guidelines for research and an integration of findings*. New York: Pergamon Press.
- EKMAN, P. (2003). «The Physiology of Facial Expressions». <http://discovermagazine.com/2005/jan/physiology-of-facial-expressions/article_view?b_start:int=0&-C=> [consulted: 28/01/2011].

- EKMAN, P.; FRIESEN, W. V. (2003). *Unmasking the face*. Cambridge, MA: Malor Books.
- EKMAN, Paul (2004). *Emotions revealed understanding faces and feelings*. London: Phoenix.
- FRESNO, Nazaret (forthcoming). «Experimenting with characters. An empirical approach to the audio description of real characters». To appear in Remael, A.; Carroll, M. and Orero, P. *Media for All: The Audiovisual Turn*. Amsterdam: Rodopi.
- FRYER, L. (2010). «Audio description as audio drama – a practitioner’s point of view». *Perspectives*, 18: 3, p. 205-213.
- FAIGIN, G. (2003). *The Artist’s Complete Guide to Facial Expression*. New York: Watson-Guptill Publications.
- HOLŠÁNOVÁ, J. (2008). *Discourse, Vision and Cognition*. Amsterdam: John Benjamins Publishing Company.
- IGAREDA, Paula (forthcoming). «The audio description of emotions and gestures in Spanish-spoken films». To appear in Serban, A.; Lavour J.-M. and Matamala, A. (eds.). *Audio-visual translation in close-up: practical and theoretical approaches*. Bern: Peter Lang.
- KRESS, G. (2003). *Literacy in the New Media Age*. Oxon: Routledge.
- LEHMAN, P.; LUHR, W. (1999). *Thinking about movies: Watching, questioning, enjoying*. London: Harcourt and Brace.
- MARCHANT, P. *et al.* (2009). «Are you seeing what I’m seeing? An eye-tracking evaluation of dynamic scenes». *Digital Creativity*, 20: 3, p. 153-163.
- MASCARENHAS, R. (2009). Audiodescription and discursive strategies: A study on the construction of the characters in «Little Miss Sunshine». Paper presented at the Media for All 3 conference. Antwerp, October 22nd-24th.
- MITAL, K. P.; SMITH, T. J.; HILL, R.L.; HENDERSON, J.M. (forthcoming). «Clustering of Gaze During Dynamic Scene Viewing is Predicted by Motion». *Cognitive Computation*.
- MONACO, J. (1977). *How to Read a Film*. 4th edition 2009. Oxford: Oxford University Press.
- ORERO, P. (forthcoming). «Film Reading for Writing Audio Descriptions: A Word is Worth a Thousand Images?». To appear in Taylor, C. & Perego, E. (eds.).
- ORERO, P.; VILARÓ, A. (forthcoming). «Eye tracking Analysis of Minor Details in Films for Audio Description». To appear in Perego, E. & d’Ydewalle, G. (eds.). *Broadening the horizons in AVT. Eye-tracking-based studies on the perception and accessibility of subtitled films*. Amsterdam: Benjamins.
- PORTER ABBOTT, H. (2002). *The Cambridge Introduction to Narrative*. Cambridge: Cambridge University Press.
- PRINCE, S. (1993). «The Discourse of Pictures: Iconicity and Film Studies». *Film Quarterly*, 47: 1, p. 16-28.
- REMAEL, A. (forthcoming). «For the use of sound. Audio description and film sound – a few key issues and a brief case study». To appear in Braun, S. & Rodgers, M. (eds.).
- TIAN, Y.; KANADE, T.; COHN, J. (2001). «Recognizing Action Units for Facial Expression Analysis». *IEEE Transactions of Pattern, Analysis and Machine Intelligence* 23: 2 p. 97-115.

Filmography

- LASSETER, J. (1989). *Knick Knack*, Pixar, US.
- (2006). *Cars*, Pixar, US.
- STANTON, A. (2008). *Wall-E*, Walt Disney, US.
- ZEMECKIS, R. (2000). *Castaway*, 20th Century Fox, US.