EYE TRACKING ANALYSIS OF MINOR DETAILS IN FILMS FOR AUDIO DESCRIPTION[1]

Pilar Orero

Universitat Autònoma de Barcelona (Spain)
pilar.orero@uab.cat

Anna Vilaró

Universitat Autònoma de Barcelona (Spain)
CAIAC (Centre d'Accessibilitat i Intel·ligència Ambiental de Catalunya)
anna.vilaro@uab.cat

Abstract

This article focuses on the many instances when minute details found in feature films may have direct implications upon the development of both the visual and plot narratives. The main question we would like to ask examines whether very subtle details which may easily go unnoticed by the viewer should be audio described. To assess the visual consciousness of such minute details, a perception experiment was conducted using eye-tracking technology and questionnaires. Though the result is not conclusive, it shows a clear methodological approach in the field of the audio description of visual details, and does give some indication as to what should be taken into consideration in future studies and analysis. The article concludes by hinting at further tests and analyses which could be undertaken using eye-tracking technology.

Resumen

Este artículo estudia los numerosos pequeños detalles que hay en la narrativa visual de las películas y que aunque aparentemente insignificantes pueden tener una

1. This research is supported by the grant from the Spanish Ministry of Science and Innovation FFI2009-08027, Subtitling for the Deaf and Hard of Hearing and Audio Description: objective tests and future plans; the Catalan Government funds 2009SGR700; and the European Project AD LAB: Audio Description. Lifelong Access for the Blind with reference no. 517992-LLP-1-2011-1-IT-ERASMUS-ECUE.

repercusión directa en el desarrollo de la narrativa visual y la trama. El tema principal es analizar si los detalles muy sutiles –que fácilmente pueden pasar desapercibidos para el espectador– deben describirse. Para evaluar la conciencia visual de estos detalles, hemos llevado a cabo un experimento utilizando la tecnología de *eye-tracking* acompañada de cuestionarios. Aunque el resultado no es concluyente, debido a la complejidad del formato del corpus de análisis, se muestra un claro enfoque metodológico en el campo de la audiodescripción de los detalles visuales, y se apunta a futuros estudios y análisis. El artículo concluye con una alusión a otras pruebas y análisis que podrían llevarse a cabo utilizando la tecnología de *eye-tracking*.

Keywords: Audio description. Eye tracking. Media Accessibility. Audiovisual Translation.

Palabras clave: Audiodescripción. *Eye tracking*. Accesibilidad a los medios. Traducción audiovisual.

Manuscript received on June 26, 2011; Definitely accepted on November 15, 2011.

1. Introduction

Andrew Holland, who has been audio describing for theatre in the UK for many years, defines audio description (AD) as "a way of translating artistic material from one medium to another [...] [that] should aim to get to the heart of a work of art and to recreate an experience of that work by bringing it to life" (2009: 184). AD is an access service which provides an *ad hoc* narrative to any artistic representation, from a surrealist painting to a circus act. AD can trace its origins back to the rhetoric figure of ekphrasis, where "graphic and often dramatic description of a painting, a relief or other work of art is provided" (Pujol & Orero 2007: 49) or hypotyposis, defined by Eco (2003: 104) as "the rhetorical effect by which words succeed in rendering a visual scene". AD has been considered a type of translation: intersemiotic, intermodal or cross-modal (Benecke 2007; Bourne & Jiménez Hurtado 2007; Braun 2007 & 2008; Orero 2005). Both audio description *per se* and eye-tracking analysis applied to audiovisual translation and media accessibility are relatively new avenues of research. Audio description has been studied from a descriptive perspective, departing from the recommendations and guidelines drafted by its practitioners – from their own professional experience. However, these recommendations and guidelines were not written with any scientific basis. New experimental studies taking into consideration eye-tracking technology have started to provide very interesting insights for both researchers and practitioners. It is hoped that articles and studies which take this methodological approach will help to improve decision making when drafting AD scripts, which up until now has been based simply on personal choice and common sense.

2. "Describe what you see": Some basic concepts of visual perception

Some basic concepts need to be addressed with regard to audio description before embarking on experiments in perception and cognition, particularly when the research approach is a multidisciplinary one.

Even after watching the same film, different people have different recollections and interpretations, and in some cases some details are observed by

some while going unnoticed by others. How can perception be so different? This issue recurs time and again throughout this article.

A question posed since antiquity is how the exterior world travels into the interior self (Goldstein 2006). A century ago, the same enquiry gave rise to the field of psychology as a scientific discipline, which examines the limits and characteristics of human perception.

From an evolutionary perspective, it can be said that our senses are designed to help us survive (Gibson 1986). In this sense, the recurrent metaphor of the photographic camera to explain human visual perception is limited exclusively to optics and eye function. While it is true that the eye works as a *camera obscura* (the basis of cameras), the camera will never be able to reveal the content of the pictures taken (Maiche & Gómez, forthcoming). On the other hand, for humans, it is actually a trivial activity to identify and enumerate the objects in our visual range (Holland 2009) – this is what some AD guidelines and studies advocate: describe what you see (Snyder 2008).

Proximal stimuli are often ambiguous, since retinas are flat surfaces where a tri-dimensional world is reflected with depth. This allows for the same proximal stimuli to correspond to different objects in real life, i.e. a pencil held at a few centimetres from an eye may generate the same stimuli as a tree trunk 100 metres away. Hence, compared with the camera, we do not simply register pictures. Nowadays we know that our perception is far from a simple reflection of the world but it is instead an *interpretation*. We interpret from previous knowledge and experiences stored in our memory, as well as from many other factors, such as our emotional state, cultural context, personal expectations, etc. See for example the work of Purves, Lotto and Nundy relating to visual illusions (2002). They propose that what we see is what the stimulus signified to us in the past, indicated by behavioural experiences of success or failure, rather than what it actually is in the present. Thus, in order to perceive we need some knowledge which will allow us to untangle these situations through the generation of a concrete situational interpretative hypothesis. In short, for each case we don't perceive solely what is gathered by the retina, but what the brain establishes according to an interpretative hypothesis. This mechanism provides complementary information to that provided by light, allowing for a better environmental adaptation and hence a higher chance of survival (Gibson 1986).

During our everyday exploration of the world, and whilst we are generally unaware of it, our eyes jump from one point to another. These fast eye movements are called saccades. Eye movements are important since humans only have high visual accuracy in the fovea, which is a tiny region of the retina.

Through eye movements, we can direct the fovea towards those areas of interest within the visual field. During a saccadic movement there is no visual perception, i.e. the visual system acts as a corrector whose objective is to maintain a stable perception of time and space through what is called saccadic suppression (Burr, Morrone & Ross 1996). It is possible to check this mechanism; if we look into a mirror from one eye to the other, we will never see the eye-movement. Without this mechanism, vision would be reminiscent of a blurred film. Between two saccades there are fixations: periods of relative stillness that allow the eye to look clearly at the chosen focused area. When looking, we don't capture an instant detailed picture of the surrounding world, but instead capture the most relevant information through a perceptive strategy.

Eye movements are directed by visual attention and this plays a key role in the active process of visual perception. It can be said that visual perception (and perception in general) has a sensorymotor aspect (O' Regan & Noë 2001). Personal experience and expectations provide a large amount of the information needed to provide meaning to the flux of visual data input from the retina. For this, and indeed in every moment, an active exploratory process of the visual range is guided by cognitive needs. Alfred Yarbus (1967) showed how eye movements are revealing with regards to one's exploratory strategies when looking at a picture. In his experiments Yarbus presented his subjects with a set of images in order to study patterns of eye movement. Figure 1 shows different patterns of eye movement on the same picture when subjects were required to provide different information about the image: 1. Free examination; 2. Estimate material circumstances of the family; 3. Give the ages of the people; 4. Surmise what the family had been doing before the arrival of the unexpected visitor; 5. Remember the clothes worn by the people; 6. Remember positions of the people and objects in the room; 7. Estimate how long the visitor had been away from the family.

In figure 1, there are some straight thin lines which represent fast and ballistic eye-movements while the subject explores a scene: the saccades. The densest points along the lines show fixations. Yarbus's results show how the eyes change fixation points in order to select the features of interest in a given scene. These results show how psychological factors drive the perception process. Likewise, when watching a film, different expectations, interpretations and motivational states influence which information is selected and able to be assimilated. We can conclude that our perception guides actions or behaviours to guarantee personal survival: visual perception is an active process which constructs environmental knowledge regarding our world.

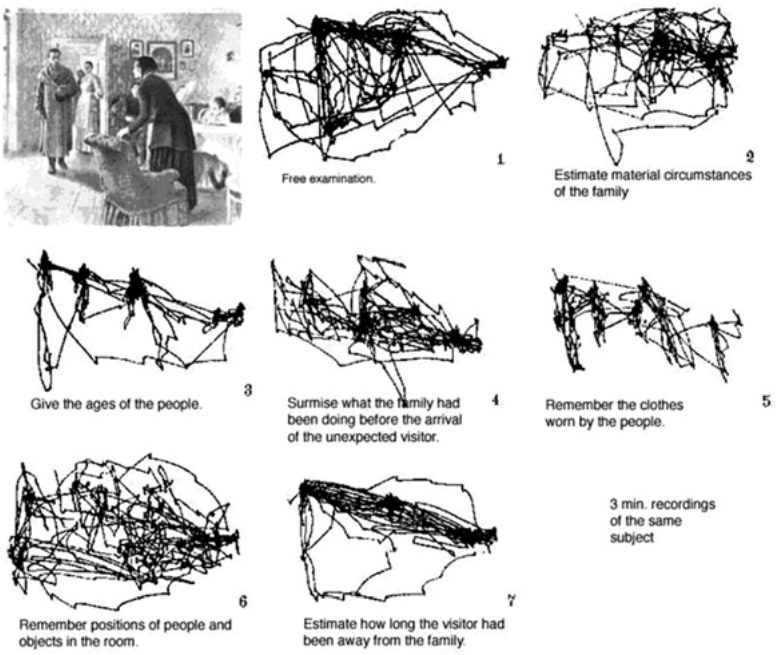

Figure 1. Eye-movement patterns of a subject required to perform various attentional tasks (Yarbus 1967: 190)

From this basic introduction to perception and cognition, we return to the original question of the different perceptions of the same film. When a film director decides to build up a scene with the addition of elements such as props and specific character traits, the intention is to convey to the audience a certain message which is enriched by the audiovisual experience. Even minute details are there for a reason, and they are often vocalized in the director's comments on DVD extras, when he explains exactly how each scene was conceived and filmed. However, when our eyes look at a film scene in an active way, searching only for those elements which are relevant, it is possible to appreciate only a proportion of these minute details.

3. What should be audio described?

There is agreement in most AD guidelines that continuity errors, mistakes, and other flaws (examples of which are shown in figures 2 and 3) should not be audio described since they don't contribute to the enjoyment of the film.

Figure 2. Left and right hand holding the cup in the same scene.

Figure 3. Moe drinks two different beers in the same scene.

But what should be done with elements which are a small part of the characterization and are only shown very briefly? We are referring to frames which are hardly ever noticed the first time a viewer watches a film, but become unavoidable as soon as they are spotted. After analysing a large corpus of films (cf. paragraph below), we noticed that in some instances such details have been described whilst in others they have been avoided.

Audio description faces the unavoidable constraint of time, and hence descriptions tend to prioritise crucial and obvious information. Nevertheless, even though in some films directors include specific information which may not be relevant to the development of the story, these can be enjoyed by those who pay close attention to minor visual elements, and either contribute to the development of the plot or help to build the traits of a character. Although these brief flashes of extra information may not seem to have a clear function, they have been included for a reason, and it is arguable whether audio description should omit such interesting frames. This issue has been raised time and again particularly following when an excerpt of Woody Allen's *Match Point* (2005, USA) was shown as an example of one of the first commercial AD releases in Spanish (see figure 7). In the Spanish AD, Ian (Nola's neighbour) is described as "un hombre negro de treinta y tantos años baja la escalera" ('A thirty-some black man comes down the stairs'). Is the colour of his skin a relevant detail contributing to the viewer's enjoyment of the story? To give a second example, why in the TV series *The No 1 Ladies' Detective Agency* (Minguella 2008-9, BBC, UK/USA; see figure 4) is there no mention of the skin colour of any of the actors? Arguably, when most of the actors in a film are white, those who are black are there for a purpose (and vice versa), since the colour of the skin may have implications in terms of the portrayal of character traits, or perhaps the colour of the person's skin is intended to make some sort of unsaid political statement. But in the AD of films such as Michael Gondry's *Be Kind Rewind* (2008, USA), Mike, one of the two main characters, is described as "black". In fact, many of the other actors are black, and the neighbourhood where the film takes place is a black district of Passaic, New Jersey, where the jazz pianist Fats Waller lived: an important factor in the development of the story. The other main character, Jerry Gerber, is white, yet that is never mentioned. The owner of the shop, along with many other characters, is also black, and yet this is also never mentioned.

There are further examples of this irregular description of ethnic origin, such as in the opening credits of *Rat Race* (2002, USA), directed by Jerry Zucker, where Cuba Gooding Jr is audio described as "a black", while Whoopy Goldberg's skin colour is never mentioned in the audio description.

Away from the issue of skin colour, there is an example from the Spanish AD of the film *Torrente 3* (Santiago Segura, 2005, Spain), where the description explains how the main character, Torrente, uses two empty tins of pickles and olives to lift weights: "Sentado en la cama de su piso bebe güisqui y levanta pesas. Las pequeñas son dos latas de banderillas, las grandes, de aceitunas" ('In his flat, sitting on his bed, he drinks whisky and lifts some weights:

Figure 4. Frame from *The No 1 Ladies' Detective Agency* (Minguella 2008-9).

the little ones are tins of pickles, the large ones are tins of olives'). Why is this information given when there are so many other details (see figure 5) to choose from in Torrente's derelict flat which may have given a wider overall impression of the slum he inhabits?

How much should be described when there is a strict time restriction is a question to which there is no straight answer. However, this has already been raised by, for instance, Vercauteren (2007), and also by Braun (2008: 21), who claims that "more research is required into audience expectations with regard to type and amount of information in the descriptions". Data from the study by Bourne and Jiménez Hurtado (2007: 177) of the Spanish and English audio-described versions of Stephen Daldry's film *The Hours* (2002, USA/UK) show the clashing approaches of two AD traditions, with a telling

Figure 5. Excerpt 1 – *Torrente 3.* Torrente, the main character of the film, is lifting some weights in a fitness routine.

difference in the AD length of almost 3,000 words: more than 7,000 words in English and almost 5,000 words in Spanish. This *décalage* comes as no surprise considering the describer responsible for the Spanish AD tradition: Javier Navarrete.

Navarrete himself (2005) states that the Spanish audience does not require as much information as is provided in an English AD. Navarrete's succinct AD style is confirmed by Bourne and Jiménez Hurtado (2007), who point out that there was "comparatively little information in Spanish concerning character and setting", i.e. there was relatively scant "detail with regards to clothes, expressions and situational context". This is a very interesting issue which needs further research, since information provided by an AD may reach the point of saturation where an audience can neither process nor remember any further details. This is the case, for example, in the animation *Monsters Inc,* directed by Docter and Silverman (2001, USA), where the characters have the following uninterrupted AD:

> J.P. Sullivan is a huge shambly monster covered in bluey-green fur with purple patches. He has lots of teeth, little ears and small curved horns on his head. His short, thick tail drags on the ground. Mike Wazowski is much smaller than James, his body and head is just one green circular blob with a single huge eye above his mouth. He has two tiny horns and two very thin green legs. Mr Watermoose is a fat monster in a jacket, red waistcoat and a bow-tie. He has five eyes, several double-chins and walks on six long thick crab's legs. Randall is a reptilian monster with a wide fat head and a big mouth full of sharp teeth. Three long sharp antenae drip from the back of his head. He has long slinky body and like J.P. Sullivan he walks upright. He can also disappear by taking in the colour of his surroundings. Celia is slim with

> one leg and one eye, her hair consists of five live rattle snakes. Roz is big, green, has a wedge of grey hair and wears swept up glasses and badly applied lipstick. All the scary monsters have smaller monsters who help them, like Mike and J.P. Sullivan.

This case is quite representative of the issue, since the target audience is children and arguably the AD offered can be fully understood and remembered by young users (Orero, forthcoming). In order to begin studying what and how much should be audio described and how to discriminate between the information, an experiment was set up involving some perception tests carried out using eye-tracking methodology and complementary questionnaires.

4. Assessing the visual input in films

While watching a film, the viewer takes as his one goal the arranging of events. This sorting can be done with the help of cues gathered from the many elements present in a film narration (Bordwell 1985). However, while we may perceive several sensorial inputs simultaneously, for example being aware of the colour of the tiles in the toilet whilst realizing that the toilet paper is finished (Tallis 2008), we can only *focus* on small pieces of information and express one idea at a time (Chafe 1980). Therefore, although we can perceive many simultaneous visual and auditory inputs, we can only express one of these at a time. Which one will be chosen when performing an audio description? Trying to understand what information should be prioritised when drafting an AD led us to design and set up the following experiment.

4.1 Objective

We have designed our experiment departing from Holsanova's (2008) assertion that verbal and visual foci are the two windows to the mind, and that they are the basic components that need to be studied in order to understand the cognitive process of the reception of an audiovisual product, and the epigonic activity of producing a narrative text. On one hand, verbal focus is "the basic discourse-structuring element that contains activated information" (Holsanova 2008: 6). This information is considered a central element in the speech unit and it usually takes the form of a sentence or a clause. In their research, both Chafe (1980) and Holsanova (1996, 2008) commence by using data gathered from spoken utterances, and we have also prepared questionnaires which will elicit oral comments. On the other hand, verbal focus is a unit which, according to Chafe (1980), coincides with one idea, taking into consideration the fact that the human mind can only formulate one idea at a

time. This fits with the departing hypothesis of Relevance Theory by Sperber and Wilson (1986) as applied to audio description. When there are multiple pieces of semiotic information available, which is the prominent single idea that is chosen to be audio described?

In this study we want to test if the AD of minute details offered in films matches the eye gaze and its intensity. There are many elements in a film which we could have chosen to create the material for our experiment, but it was decided to test the most important element in film narrative: the character (Bordwell 1985). In order to narrow down which features to test from the character, we examined the three different character components as suggested by Phelan (1996: 216): the synthetic, the thematic, and the mimetic (see section 3.3).

Prior to testing the correlation between AD and the perception of minute details, it is necessary to check what is perceived and what is processed. That is, how far people's perception of the same visual input will produce uniform perception and cognitive experiences, as has been examined by Romero-Fresco (forthcoming) regarding when people watch television news. Interestingly, Romero-Fresco shows how the perception of subjects exposed to the same excerpt of news varies considerably to the extent that some participants stated that Tony Blair was present in a news clip, an assertion which was in fact untrue.

4.2 Experiment Design

Following Holsanova's (2008) experiments testing both the verbal and visual foci, we decided to embark on an experiment where visual input would be measured through eye-tracking technology and questionnaires.

Close attention was paid to the choice of video segments, as we had to find representative elements condensed into a very short time span. As explained by Germeis and d'Ydewalle (2007: 458), "films consist of a series of shots edited together to make a coherent visual story. A shot is a single run of the camera, while a cut is the transition between the end of one shot and the beginning of the next." The segments used were part of a shot offering ample information. In a sense one could say that the shots were saturated with information and detail, since in a two-minute clip the element which is the object of analysis is shown without any preamble. The choice of clips was governed by the issues which we wanted to test, in this case the three different components of a character: synthetic, thematic and mimetic –explained below–.

4.3 The corpus

With all these considerations as the basis for the research, data was gathered through user tests examining the perception of sensorial stimuli in the form of three different excerpts, each chosen for a different reason (see figures 5, 6 and 7). The films used were *Torrente 3* (S. Segura, 2005, Spain), *Raising Helen* (Garry Marshall, 2004) and *Match Point* (W. Allen, 2005, USA). The clips lasted 47 seconds, 54 seconds and 34 seconds, respectively and were shown without any audio in order to avoid exogenous attention control.

In the classification of characters by Phelan (1996: 216), we find the s*ynthetic* – a character that "plays a specific role in the construction of narrative as made object". Torrente, portrayed in the first excerpt (see figure 5), has been chosen as a representative of this type.

It is important to note all the details which surround Torrente: his clothes, his haircut, his moustache, etc., and also the many elements which are present in his flat. This extreme characterisation is a significant part of the film's narrative which, after all, is the portrayal of an archetypal Spanish fascist in 21st century Spain. Torrente is a right-wing Francoist ex-cop; he is selfish, petty, chauvinistic, misogynistic, racist and homophobic. His attitudes towards disability, religion and sex are as politically incorrect as humanly possible. This clip is chosen in order to check if the audience looks at the weights he is lifting, since they are very 'do-it-yourself' made as they are from recycled cans of pickles and olives regularly sold in bars and restaurants. The film *Torrente 3* was audio described in Spanish in 2005, and precise information regarding the weights was provided: "Sentado en la cama de su piso bebe güisqui y levanta pesas. Las pequeñas son dos latas de banderillas, las grandes, de aceitunas" ('In his flat, sitting on his bed he drinks whisky and lifts some weights: the little ones are tins of pickles, the large ones are tins of olives'). The key question here was how many people would notice the cans of olives and pickles without the exogenous attention control of the audio description?

The second excerpt was taken from *Raising Helen*, and was chosen to represent the *thematic* character, defined as "any character representative of one or more groups and functions in one way or another to advance the narrative's thematic concern" (Phelan 1996: 216). As can be seen in figure 6, two famous actresses are seated in the first row of a fashion show: Sofia Loren and Paris Hilton. They are part of a crowd of people watching a fashion show. The AD that is presented simultaneously to the selected scene read "Two male and two female models strut on to the stage wearing striking underwear. Sara in a pink stripped cardie spots Amber and her dog."

Figure 6. Excerpt 2 – Paris Hilton and Sofia Loren in *Raising Helen*.

The clip was selected to examine through eye tracking whether participants looked at both women, or if instead they fixed their attention on only one of them and, if this was the case, which one. The key point here is that human faces are the priority area of interest when displayed on a screen (Palermo & Rhodes 2007). In this case there is a crowd, and in that crowd there are two people of particular interest. Should the audio description mention there is a crowd of people including two famous actresses, or instead simply limit itself to describing a crowd watching a fashion show? The element of surprise was also present since this clip involves what can be considered as a cameo appearance by Sofia Loren and Paris Hilton.

Finally, in excerpt 3 (figure 7) taken from *Match Point,* we meet Nola's neighbour Ian, a young black man. In this case the clip was chosen to represent the *mimetic* character from Phelan's classification (ibid.), defined as a "character like a person" in the sense that they imitate real life and have the appearance of normality. In this case, the subject is the person next door. Woody Allen's choice of a young black person as the neighbour has the effect of normalising the status of black people, deliberately contradicting any existing prejudices or stereotypes held by the viewer. The neighbour is well-dressed, well-spoken and is shown to be a kind and caring person since, in only a very brief appearance, he both knocks on an old lady's door to see if she needs anything from the shops and also checks with Nola details of the walkman she wanted to buy. Viewers, according to Bordwell (1985: 30), are humans with many prejudices and personal perspectives, the "ideal viewer" being an impossible reality. Thus our interpretation of this character, and the

Figure 7. Excerpt 3 – *Match Point*. In the picture, Ian, who is Nola's neighbour.

excerpt was included in order to check if participants noticed the fact that he is black. Furthermore, should the audio description highlight this fact? The AD says "Un hombre negro de treinta y tantos años baja la escalera" ('A thirty-something black man comes down the stairs').

4.4 Participants

Eighteen sighted volunteers were recruited and took part in the test, comprising 10 male and 8 female. They were all university students and their average age was 29.5. Participants had normal or corrected vision.

4.5 Apparatus

Tests were performed on a Tobii T-60 eye tracker with a 60 Hz sampling rate and a 1280×1024 display. The recording had 0.5 accuracy, about 10×10 pixels at 50 cm distance from the observer.

4.6 Procedure

The test was set up to consist of two phases. The first phase involved the recording of the gaze whilst watching the excerpts. To achieve this, volunteers were sat centered in front of the eye tracking display and asked to adopt a comfortable position, keeping movements to a minimum. All recordings were calibrated in the eye tracker in order to guarantee good quality data. Participants were told that they would see three excerpts from three different films. They were also told to watch the three excerpts as if they were at home sitting

in front of the television. The first phase was approximately three minutes in duration. The excerpts were presented sequentially always in the same order (*Torrente 3, Raising Helen,* then *Match Point*), and were separated by a brief pause of a blank screen. After watching the excerpts, in the second phase volunteers were asked about the target element in each case. The questions were:

1) What does Torrente use as weights?
2) Are there any famous actresses in the crowd?
3) Was Nola's neighbour white or black?

To control any possible memory bias, the participants were also asked if they had previously seen any of the films.

4.7 Results

All participants were accepted, since recordings were correct and declared valid. Eye fixations data were obtained for the areas of interest (AOI) for each excerpt and heat map visualizations were created. AOI were defined prior to the experiment by the target element we wanted to test according to visual details.

Heat maps of all the recordings are presented below. Areas with a higher number of fixations (counts) during the scene are coloured in red. This indicates where participants tended to look. Heat maps were created from data of a very brief scene. The length of each segment calculated is mentioned in each case.

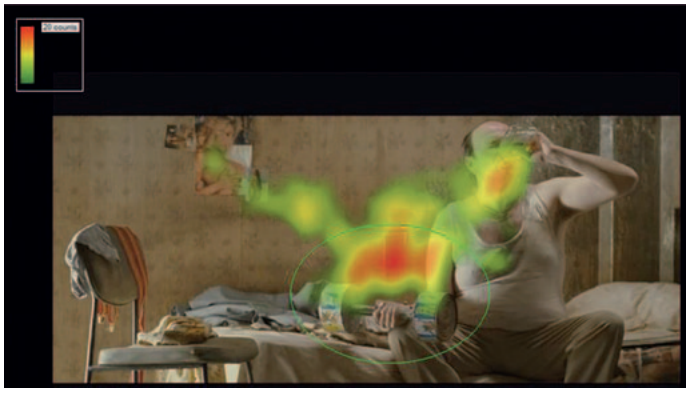

Figure 8. Excerpt 1 – Heat map for *Torrente 3*.

Figure 9. Excerpt 2 – Heat map for *Raising Helen*.

Figure 10. Excerpt 3 – Heat map for *Match Point*.

Data for excerpt 1 is from a 15-second scene. The main fixation length in the AOI is 4.35 seconds (standard deviation = 1.1 seconds). The heat map in figure 8 shows the accumulated number of fixations from all the participants.

Data from excerpt 2 is based on an 11-second scene with two AOI: the faces of Paris Hilton and Sofia Loren. The two AOI comprise the same area of the total image. A fixation of 4.6 seconds (standard deviation = 3.1) has been observed on Paris Hilton's face, and a shorter period of 1.2 seconds (standard

deviation = 1) on Sofia Loren's face. The heat map in figure 9 shows the accumulated number of fixations.

Data from excerpt 3 is from a 4-second scene. The main fixation length in the AOI is 2.1 seconds (standard deviation = 1.3 seconds). The heat map in Figure 10 shows the accumulated number of fixations.

As can be observed from the images, all the target features have fixations. If we look at individual data, the same fixation pattern is repeated between participants and it could be said that the scan patterns for each excerpt are very similar for the different participants. Regarding the analysis of the second phase of the test (the stage that involved answering questions), for excerpt 1, 33% realized that the weights were made from cans; regarding the second excerpt, 44% recognized the famous actresses in the fashion-show; finally, 83% answered that the man going downstairs was black.

5. Discussion

Since perception is selective (i.e. we cannot simultaneously perceive everything around us), attention mechanisms take place with the function of selecting one of the many information inputs in a given moment. Attention can be either voluntary (endogenous control) or automatic (exogenous control).

Results from the eye tracker show that eye fixations and scanpaths are very similar among all participants, suggesting a role for exogenous control of attention. In 1980, Just and Carpenter formulated the Eye-Mind Hypothesis, according to which there is no perceptible lag between what is fixated and what is processed. This hypothesis was often questioned in light of *covert attention* (Posner 1980), meaning the attention to something that one is not looking at. Thus, eye tracking recordings would often show not where the attention had been, but only where the eye had been looking, meaning that eye tracking would not necessarily indicate cognitive processing. It is possible that when a participant fixed his eyes on a part of the scene (e.g. on Torrente's weights or Sofia Loren's face), his attention wasn't actually there but instead was actually somewhere else (such as on Torrente's wall poster or Paris Hilton's bust). Consequently, the part of the scene where the participant happened to fix his eyes would not necessarily have been consciously perceived. In this sense, we still cannot directly infer specific cognitive processes from a fixation on a particular object in a scene. At present, the consensus is that visual attention is always slightly (by between 100 and 250 milliseconds) ahead of the eye (Hoffman 1998), but as soon as attention moves to a new position, the eyes will want to follow (Deubel & Schneider 1996).

With eye tracking measurements, we added one more testing procedure to assess the conscious perception: answering a questionnaire. Data gathered from the second phase of the experiment showed how perception is not uniform amongst all participants. Another issue which should be taken into consideration when interpreting data is the participant's own memory. It is possible that there was some biased perception for this reason. 72% of the participants had seen *Torrente 3*, 1% had seen *Raising Helen*, and 5% had seen *Match Point*. Given the varied prior experience the participants had of the different films, the impact of this is difficult to assess. That said, it should also be noted that tests were performed in the summer of 2009, with the films having been released in 2004 and 2005. The ideal situation would have been that no participant had previously seen any of the films.

Regarding our results, this implies that fixation on a specific element does not automatically mean that information is processed and remembered. It can be said that covert attention is implied, since there isn't agreement between eye-tracker data and questionnaire data. For this reason, issues such as movement in the film as AOI should be studied, or at least taken into consideration, such as the movement of Torrente's weights. Is attention focussed on the weights because they provide essential clues to the development of the character and the plot of the film, or are the results obtained due to the eye scanning for movement? Matching these two sources of information allows us to observe the disagreement between fixations and perception, and this will be the starting point of a future article.

6. Conclusions

Some recurrent audio description research areas have been analysed and, given their importance, it is clear that they need to be studied further. Take for example the theory that perception stays at the level of the discourse or text type known as description, defined by Chatman (1990: 3) as "the evocation of the properties of objects for their own sake". This essentially means that description performs the same function as a camera. Perception is the sensorial stage before cognition which in turn would correspond to the text type known as narration. It would be very interesting to proceed along this line of research in order to evaluate the different text types (description vs. narration) and the function of each according to cognitive requirements, taking into account the point made by Chatman (1990: 2), "the old prejudice that Narrative somehow dominates Description. Most texts utilize one overriding text-type, but this is generally subserved by other text-types".

Regarding the data obtained from the three-clip analysis, a quick comparison was established with the interest shown on some spots by sighted viewers and the existing commercial ADs available on DVD. In all three cases there was no correlation between what was audio described in the commercial DVD and the areas of interests identified by the study. Again, this correspondence shows the deviation from the information offered in the AD and the visual clues, used by sighted viewers to construct the lineal narrative. The disparity of information may correspond to what is a recurrent topic in AD, namely "subjectivity", a much maligned term in AD, though unavoidable since we are in the realm of creative writing. We also need to take into consideration the relevance of information according to each person who watches and then drafts the AD, hence while in Translation Studies the term "subjective" translation has never been an issue, perhaps we could move away from this tendency in the field of research in AD. The reality of AD scripts as human constructs, with all its possible interpretations is a fact. Research in this issue should take more interesting directions such as understanding basic narratological elements and their function, as in the work of Kruger (2009 & 2010) and Vercauteren (forthcoming).

Finally, given the lack of a critical mass regarding audio description, we would like to highlight some points which should be given due consideration when embarking on this type of research. Close attention was paid to the choice of video segments (Germeis & d'Ydewalle 2007). The segments used were part of a shot offering ample information. Little attention was paid to the perceptual processing of film segments (Hochberg 1986; Hochberg & Brooks 1996), since the element under investigation was extracted from the environment of a cut. That said, the level and intensity of attention required when watching a two-minute segment without audio and a ninety-minute film must vary considerably. Can evaluation performed under such different circumstances still count as a valid analysis of perception, attention, memory, and cognition?

Regarding the interpretation of data, a key consideration is the fact that subjects were not watching the clips in a natural and relaxed environment. An ideal situation would be to conduct tests without the knowledge of the participants. How a relaxing and cosy atmosphere can be achieved in user labs, even if only still offering a *simulation of* home conditions, is an interesting challenge.

To conclude, once the technology is understood and the range of its use properly evaluated, it may well be that eye tracking could prove to be an efficient tool for the analysis of audio description. The new Special Issue from

Perspectives 2012, edited by Jan Louis Kruger and Iwona Mazur, where many articles take this experimental approach, highlights the level of interest in the methodology. While the technology is nothing more than a piece of hardware, with some detailed studies and careful methodology it may be possible to make progress in the field and obtain meaningful data which will be of great help when investigating some AD theories.

References

BENECKE, Bernd. (2007) "Audio Description: Phenomena of Information Sequencing". *MUTRA*. Full-text version at: <http://www.euroconferences.info/proceedings/2007_Proceedings/2007_Benecke_Bernd.pdf> [retrieved on 20/05/2009].

BORDWELL, David. (1985) *Narration in the Fiction Film*. London: Methuen.

BOURNE, Julian & Catalina Jiménez Hurtado. (2007) "From the Visual to the Verbal in Two Languages: a Contrastive Analysis of the Audio Description of *The Hours* in English and Spanish". In: Díaz-Cintas, Jorge; Pilar Orero & Aline Remael (eds.) 2007. *Media for All: Subtitling for the Deaf, Audio Description and Sign Language*. Amsterdam: Rodopi. pp. 175-188.

BRAUN, Sabine. (2007) "Audio Description from a discourse perspective: a socially relevant framework for research and training". *LANS* 6. pp. 357-372.

BRAUN, Sabine. (2008) "Audiodescription Research: state of the art and beyond". *Translation Studies in the New Millennium* 6. pp. 14-30.

BURR, David; Maria Concetta Morrone & John Ross. (1996) "Selective suppression of the magnocellular visual pathway during saccades". *Behavioral Brain Research* 80. pp. 1-8.

CHAFE, Wallace. (ed.) (1980) *The Pear Stories: Cognitive, Cultural and Linguistic Aspects of Narrative Production*. Norwood & New Jersey: Ablex.

CHATMAN, Seymour. (1990) *Coming to Terms. The Rhetoric of Narrative in Fiction and Film*. Ithaca & London: Cornell University Press.

DEUBEL, Heiner & Werner Schneider. (1996) "Saccade target selection and object recognition: Evidence for a common attentional mechanism". *Vision Research* 36:12. pp. 1827-1837.

ECO, Umberto. (2003) *Mouse or Rat? Translation as Negotiation*. London: Weidenfeld & Nicolson.

GERMEIS, Filip & Géry d'Ydewalle. (2007) "The psychology of film: perceiving beyond the cut". *Psychological Research* 71. pp. 458-466.

GIBSON, James J. (1986) *The ecological approach to visual perception*. Hillsdale & New Jersey: Ed. Lawrence Erlbaum Associates.

GOLDSTEIN, E. Bruce. (2006) *Sensation and perception*. Pacific Grove, CA: Wadsworth.

HOLSANOVA, Jana. (1996) "Attention Movements in Language and Vision". In: *Representations and Processes between Vision and NL. Proceedings of the 12th European Conference of Artificial Intelligence*. pp 81-83). Budapest, Hungary. Full text-version at: <http://www.lucs.lu.se/Jana.Holsanvoa/PDF/Holsanova.1996.project.PDF>

HOLSANOVA, Jana. (2008) *Discourse, vision, and cognition*. Human Cognitive Processes 23. Amsterdam & Philadelphia: John Benjamins Publishing Company.

HOCHBERG, Julian. (1986). "Representation of motion and space in video and cinematic displays". In: Boff, Kenneth R.; Lloyd Kaufman & James P. Thomas (eds.) 1986. *Handbook of perception and human performance 1. Sensory processing and perception*. New York: Wiley.

HOCHBERG, Julian & Virginia Brooks. (1996) "The perception of motion pictures". In: Friedman, Morton P. & Edward C. Carterette (eds.) 1996. *Cognitive ecology*. San Diego, CA: Academic Press. pp. 205–292.

HOFFMAN, James. (1998) "Visual attention and eye movements". In: Pashler, Harold (ed.) 1998. *Attention. Studies in cognition*. London: Psychology Press. pp. 119-153.

HOLLAND, Andrew. (2009) "Audio Description in the Theatre and the Visual Arts: Images into Words". In: Anderman, Gunilla & Jorge Díaz-Cintas (eds.) 2009. *Audiovisual Translation. Language Transfer on Screen*. Basigtoke: Palgrave Macmillan. pp. 170-185.

JUST, Marcel Adam & Patricia A. Carpenter (1980) "A theory of reading: From eye fixations to comprehension". *Psychological Review* 87. pp. 329-354.

KRUGER, Jan Louis. (2009). "The translation of narrative fiction: impostulating the narrative origo". *Perspectives: Studies in Translatology* 17:1. pp. 15-32.

KRUGER, Jan Louis. (2010) "Audio narration: re-narrativising film". *Perspectives* 18:3. pp. 231-249.

NAVARRETE, Francisco Javier. (2005) "Sistema Audesc: el fin de los Susurros", Seminario sobre medios de comunicación sin barreras. Full text-version at: <http://www.uch.ceu.es/sinbarreras/textos/jnavarrete.htm> [retrieved on 20/05/2009]

MAICHE, Alejandro & Leonel Gómez. (forthcoming) "La Visión: de los fotorreceptores a la percepción". In: Redolar, Diego (ed.) Barcelona: Editorial UOC.

O'REGAN, J. Kevin & Alva Noë. (2001) "A sensorimotor account of vision and visual consciousness". *Behavioral and brain sciences* 24. pp. 939-1031.

ORERO, Pilar. (2005) "Audio Description: Professional Recognition, Practice and Standards in Spain". *Translation Watch Quarterly* 1. pp. 7-18.

ORERO, Pilar. (forthcoming). "Audio Description for Children: Once upon a time there was a different audio description for characters". In: Becerra, Carmen & Elena di Giovanni (eds.) *Entre texto y receptor: accesibilidad, doblaje y traducción*. Frankfurt: Peter Lang.

PALERMO, Romina & Gillian Rhodes. (2007). "Are you always on my mind? A review of how face perception and attention interact". *Neuropsychologia* 45:1. pp. 75-92.

PHELAN, James. (1996) *Narrative as Rhetoric. Technique, Audiences, Ethics, Ideology*. Columbus: Ohio State University Press.

POSNER, Michael I. (1980) "Orienting of attention". *Quarterly Journal of Experimental Psychology* 32. pp. 3-25.

PUJOL, Joaquim & Pilar Orero. (2007). "Audio description precursors: Ekphrasis, film narrators and radio journalists". *Translation Watch Quarterly* 3:2. pp. 49-60.

PURVES, Dale; R. Beau Lotto & Surajit Nundy. (2002) "Why we see what we do". *American Scientist* 90:3. pp. 236-243.

PURVES, Dale; Willian T. Wojtach & Catherine Q. Howe. (2008). "Visual illusions: An Empirical Explanation". *Scholarpedia* 3:6. pp. 3706. Full text-version at: <http://www.scholarpedia.org/article/Visual_illusions:_An_Empirical_Explanation> [retrieved on 20/05/2009].

ROMERO-FRESCO, Pablo. (forthcoming) "Standing on quicksand: Hearing viewers' comprehension and reading patterns of respoken subtitles for the news". In: Díaz-Cintas, Jorge; Anna Matamala & Josélia Neves (eds.) *Media for All 2*. Amsterdam: Rodopi.

SNYDER, Joel. (2008) "The Visual Made Verbal". In: Díaz Cintas, Jorge (ed.) 2008. *The didactics of Audiovisual Translation*. Amsterdam: John Benjamins. pp. 191-198.

SPERBER, Dan & Deirdre Wilson. (1986) *Relevance: Communication and Cognition*. Oxford: Basil Blackwell.

TALLIS, Raymond. (2008) "License my roving hands. Does neuroscience really have anything to teach us about the pleasures of reading John Done?" *Times Literary Supplement* 11/04/2008. pp. 13-15.

VERCAUTEREN, Gert. (2007) "Towards a European Guideline for Audio Description." In: Díaz Cintas, Jorge; Pilar Orero & Aline Remael (eds.) 2007. *Media for All. Accessibility in Audiovisual Translation*. Amsterdam: Rodopi. pp. 139-150.

VERCAUTEREN, Gert. (forthcoming) "A Narratological Approach to Content Selection in Audio Description. Towards a Strategy for the Description of Narratological Time". In: Agost, Rosa; Pilar Orero & Elena di Giovanni (eds.) *Multidisciplinarity in Audiovisual Translation*. *MonTI* 4.

YARBUS, Alfred L. (1967) *Eye Movements and Vision*. New York: Plenum.

BIONOTE / NOTA BIOGRÁFICA

Pilar Orero

PhD (UMIST). Works in the CAIAC Research Centre (Universitat Autònoma de Barcelona, Spain). She started the two MAs in Audiovisual Translation at UAB, and now is the director of the Online European MA in Audiovisual Translation (http://mem.uab.es/metav/). Recent publications: *Topics in Audiovisual Translation* (2004), John Benjamins. Co-editor with Jorge Díaz-Cintas and Aline Remael of *Media for All: Subtitling for the Deaf, Audio Description and Sign Language* (2007), Rodopi. Co-editor with Anna Matamala of *Listening to Subtitles: SDHoH* (2010) in Peter Lang. Co-writer with Anna Matamala and Eliana Franco of *Voice-over: An Overview* (2010) in Peter Lang. Guest editor of *TRANS* 11 and co-guest editor with J.L. Kruger of *Perspectives* on Audio Description (2010). Leader of numerous research projects funded by the Spanish and Catalan Gov. Partner of the EC project DTV4ALL (http://www.psp-dtv4all.org/). Leads TransMedia Catalonia (http://grupsderecerca.uab.cat/transmediacatalonia)

Doctora por la UMIST. Treballa al Centre de Recerca CAIAC (Universitat Autònoma de Barcelona, Espanya). Propulsora de dos màsters en Traducció Audiovisual a la UAB, i actualment directora del Màster Europeu de Traducció Audiovisual (http://mem.uab.es/metav/). Publicacions recents: *Topics in Audiovisual Translation* (2004), John Benjamins. Coeditora amb Jorge Díaz-Cintas i Aline Remael de *Media for All: Subtitling for the Deaf, Audio Description and Sign Language* (2007), Rodopi. Coeditora amb Anna Matamala de *Listening to Subtitles: SDHoH* (2010) a Peter Lang. Coautora amb Anna Matamala i Eliana Franco de *Voice-over: An Overview* (2010) a Peter Lang. Editora convidada de *TRANS* 11 i coeditora convidada amb J.L. Kruger de *Perspectives* on Audio Description (2010). Investigadora principal de nombrosos projectes de recerca finançats pel Govern espanyol i català. Partner del projecte europeu DTV4ALL (http://www.psp-dtv4all.org/). Directora del grup de recerca TransMedia Catalonia (http://grupsderecerca.uab.cat/transmediacatalonia).

Anna Vilaró

BA in Psychology from UAB (Spain) and finishing her PhD on stereoscopic audiovisual content and subtitling and verbal and visual information processing. She lectures at the European MA in Audiovisual Translation, and the Universitat Oberta de Catalonia (Spain). She is the author of numerous papers on

audiovisual translation and eye-tracking methodology. She is the manager of the LabTab at CAIAC, UAB.

Llicenciada en Psicologia per la Universitat Autònoma de Barcelona, Espanya. Actualment està realitzant el doctorat en continguts audiovisuals estereoscòpics subtitulats i processament de la informació verbal i visual. És professora del Màster Europeu de Traducció Audiovisual i consultora a la Universitat Oberta de Catalunya (Espanya). És autora de diversos articles sobre traducció audiovisual i metodologia de seguiment ocular. És la persona encarregada del LabTav al Centre de recerca CAIAC, UAB.