Perspectives: Studies in Translatology
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/rmps20

Visual and narrative priorities of the blind and non-blind: eye tracking and audio description

Elena Di Giovanni[a]

[a] Department of Human Studies, University of Macerata, Macerata, Italy

Published online: 27 Mar 2013.

To cite this article: Elena Di Giovanni (2014) Visual and narrative priorities of the blind and non-blind: eye tracking and audio description, Perspectives: Studies in Translatology, 22:1, 136-153, DOI: 10.1080/0907676X.2013.769610

To link to this article: http://dx.doi.org/10.1080/0907676X.2013.769610

Perspectives: Studies in Translatology, 2014
Vol. 22, No. 1, 136–153, http://dx.doi.org/10.1080/0907676X.2013.769610

Visual and narrative priorities of the blind and non-blind: eye tracking and audio description

Elena Di Giovanni*

Department of Human Studies, University of Macerata, Macerata, Italy

(*Received 29 May 2012; final version received 7 January 2013*)

This paper reports on a complex experiment carried out in 2010 in Italy with sighted individuals and the blind, the primary aim of which was to investigate the applicability of eye tracking research to audio description (AD). Starting from the identification of the visual priorities of sighted individuals for two, 100-second, virtually dialogue-free clips from the same film (*Tris di donne & abiti nuziali*), the project continued with the drafting of descriptions for these clips, based on the data obtained through the eye tracking tests. These eye-tracker-derived audio descriptions were presented to a group of blind individuals alongside the traditionally drafted ADs and followed by oral questionnaires. The results of this experiment have confirmed the great relevance of eye tracking research for audio description. They have also provided valuable information to be poured into the enhancement of the practice of audio description in Italy and elsewhere.

Keywords: audio description; eye tracking; blindness; reception studies; screen translation

Both eye tracking (ET) and audio description (AD) research are enjoying an unprecedented popularity within AVT. This is proof of the truly multidisciplinary essence of the field (Di Giovanni, 2009; Di Giovanni, Orero, & Agost, 2012), the great impact of technology-based research and practices and, last but not least, the increasing attention to the right to media accessibility for the visually impaired. However, audio description and eye tracking have only recently come to form a pair in research,[1] perhaps due to the difficulty in overcoming the seemingly oxymoronic relation in which they stand. How can eye tracking experiments, which aim at testing eye movements and fixations, be employed with reference to audio description, i.e. a type of intersemiotic translation conceived for the non-viewing population?

In Europe, the first inputs to challenge this oxymoronic relation have come from the European project 'Digital Television for All' (TCP IP Ref: 224994), more precisely from the team involved in research on audio description.[2] Setting off to evaluate the possibility of defining common principles for the drafting of ADs across cultures, the research team proceeded first of all to identify differences and similarities in the visual perceptions of sighted viewers in three European areas (Italy, Poland, Catalonia), as they emerged from eye tracking tests performed on short, selected clips from the film *Marie Antoinette*. The results proved extremely

*Email: elena.digiovanni@unimc.it

interesting and suggested that similarities are to be found at different levels, thus encouraging further research based on cross-cultural ET tests.

Moving along those lines, a small research team at the University of Macerata (Italy)[3] decided to develop a 'monocultural' project, aiming to draft eye tracker-derived, Italian audio descriptions and test their reception by end users. This complex experiment, carried out over an 18-month period (March 2009–October 2010), is reported in this paper. The following pages illustrate each stage of the project and focus on the results obtained through the reception tests. Thus, starting from an oxymoronic hypothesis, the empirical research here presented not only dismisses the oxymoron but goes as far as proving the effectiveness of eye tracking in audio description research, as is demonstrated by the responses of the end users.

The project: background and objectives

Preliminary reflections

As James Monaco rightly puts it in *How to read a film*, 'a film is strongly pictorial' (2000, p. 29). Even though the role of verbal language is essential in the process of meaning making, a film, its characters and its narrative development are largely determined by the images and their movement. Images set the tone of a film, they define its context and evoke its connections with one or more cinematic genres. They stimulate a sense of communion in the viewers and arouse their emotions. The interaction of images and words is indeed essential for the reception of a film, but images on their own are often more powerful in evoking a film's interpretation than words and sounds.

The nature, juxtaposition and movements of iconic signs are, therefore, determining in the process of meaning making. This process, however, requires active participation and ultimately rests with the viewers: a film offers an array of visual stimuli that, although purposefully organized by the film creators, leave it to the viewer to make interpretive choices. Along these lines, James Monaco suggests that 'observers can act: making choices from the dramatic, pictorial, narrative, musical and environmental materials that present themselves in films' (2000, p. 37). In this perspective, viewers are in fact *observers*, they perceive the visual stimuli and cognitively elaborate the film's reception.

Viewers, spectators, observers: all three nouns refer to the visual sphere. They evoke the recourse to sight as a primary sense in the decoding and receiving of a film, which strongly marginalizes and excludes the partially sighted and the blind. Thus, it seems that the less 'viewing' an individual is, the more s/he loses her right to agency in the filmic experience.

Audio description is an extremely valuable tool that allows for the restoration of agency in the enjoyment of films and audiovisual texts by the blind and partially sighted: translating iconic signs into verbal sequences and combining them with the film's soundtrack aims to provide the sensory impaired with the elements they need to form their own interpretation and reception of a film.[4] Ideally, as is often said by scholars and practitioners,[5] AD should neither provide too much nor too little information, however difficult it is to possibly define such an ideal. It is said, in general, that AD should focus on the verbalization of those visual stimuli that are essential for an understanding of the actions and interactions, without providing any interpretation and leaving it to the blind audience to form their own ideas.

The very definition of AD has been the object of endless debates by practitioners and scholars from various fields of research: how can a *juste milieu* be defined and put in practice? Is the reception and recoding of visual stimuli by sighted individuals the only possible source of data for AD? In other words, is the outcome of this twofold process, filtered through individual experiences and inextricably linked to the perception of verbal stimuli, really what a non-sighted individual needs to understand a film? Wouldn't it be possible to isolate the visual perception of sighted viewers and use this as a basis for the drafting of AD?

These and other questions have been at the core of the few experiments carried out so far with the support of eye tracking research.[6] The one reported in the following sections is among the very few which have so far managed to cover the entire cycle, which goes from eye tracking tests with sighted viewers to comparative AD tests with the blind and visually impaired. Starting off from audience research and finishing with reception research (with the audience) was indeed one of our objectives, which will be discussed in the next section.

Objectives

In a book on the appreciation of films by people with visual impairments (*Cinema e disabilità visive*), Anna Poli reflects on the fact that the 'quality' of our perception of the world is shaped by our visual system: 'eye movements and their functions are responsible for the quality of our vision of the world around us and the details which make it what it is' (2009, p. 35, my translation). Poli goes on to state that the role of the visual system is that of a compass, which provides orientation and defines trajectories. Reflecting upon these insights with reference to AD, its function and its being inevitably derivative from the individual reception of sighted viewers, we decided to move a few steps backwards to recover what we may here call the 'raw data' of visual perception, i.e. the trajectory accomplished by the eyes – driven by the visual system – before further elaboration by the brain (Rainer et al., 2009). Thus, we decided to observe the fixations and gaze control of sighted viewers to subsequently try and apply those data to the creation of an AD unmediated by individual reception and recoding processes.

Gaze control, as defined by Henderson et al. (2007, p. 539), is the process of directing the eyes through a scene in real time, in the service of ongoing perceptual, cognitive and behavioural activity. The importance of the study of gaze control in real-world scene perception has been clearly explained by the same authors, who have provided three main reasons for this:

> First, human vision is active, in the sense that fixation is directed toward task-relevant information as it is needed for ongoing visual and cognitive computations. Although this point seems obvious to the eye movement researchers, it is often overlooked in the visual perception and visual cognition literature. [. . .] Second, eye movements provide a window into the operation of selective attention. Third, because gaze is typically directed at the current focus of analysis eye movements provide an unobtrusive, sensitive, real-time behavioral index of ongoing visual and cognitive processing. (Henderson et al., 2007)

The three reasons above provide a perfect thrust to our initial hypotheses: fixation is directed toward task-relevant information, just like a compass directs straight towards relevant targets. Eye movements provide an unobtrusive, sensitive and real-time indication of the visual and cognitive processes aimed at identifying and

tracking elements which are carriers of visual information. All of this clearly confirms that no act of vision occurs per se, i.e. without a cognitive process. However, cognitive processes accompanying the visual system and its movements are manifold; they contribute to the overall work of the so-called visual brain, whose complexity is beyond the scope of this paper and the skills of its author. Interestingly, however, Millner & Goodale (2006) offer a definition of visual brain that is twofold and allows for a sort of division of competencies. They talk about 'two separate and quasi-independent visual brains' (2006, p. 1), concerned respectively with visual perception and the visual control of action. This distinction restricts the meaning of the word 'perception' to the operations of fixation and gaze shifting and it excludes more complex processes that lead to further actions (2006, p. 3). It is in this sense that we refer to visual perception in this paper, and it is precisely on this type of perception that we aimed to focus for the first part of the experiment here described.

However, from these initial hypotheses to the testing of an eye tracker-derived AD with the blind, the path appeared long and complex. To face such a complexity in the smoothest possible way, we identified a sequence of steps to be undertaken towards the completion of the project:

(1) evaluating the fixation and gaze control of a sample of Italian sighted viewers by means of tests performed on an eye tracker;
(2) analysing the resulting data and pouring them in the drafting of ET-induced AD; and
(3) testing normally produced and ET-induced audio descriptions for the same film excerpts with a group of blind individuals and giving them oral questionnaires to evaluate their comprehension and overall reception.

The next sections will describe each of the steps above.

Stage one: eye tracking tests

The first step into the development of our project implied selecting a film and, within that film, at least two clips for our ET tests. A choice was made, taking into account only films which had already been audio described in Italy and, more specifically, ones whose AD script could be easily available to us. We opted for *Tris di donne & abiti nuziali*, a 2009 film by Vincenzo Terracciano, which premiered at the Mostra del Cinema di Venezia in the same year and was audio described by SubTi Access for one or two of the official screenings at that festival. Within the overall film, we identified two clips with features to suit our experiment: they were both 100 seconds long and each clip contained virtually no dialogue (three words in the first clip and only one in the second). Moreover, each clip features the interaction of different characters.

The first sequence (Clip 1) portrays a fairly rapid series of events that are logically chained and that rely heavily on visual perception to be understood. In this excerpt, Franco (Sergio Castellitto) is chased by Matteo (Gigio Morra), to whom he owes money. The chase weaves through the streets of Napoli, along a staircase, through a shopping arcade, in a tram station and finally inside a train. The second sequence (Clip 2) is slower and more intimate, with the narrative development unfolding mainly through gestures and facial expressions. It features two women, Josephine (Martina Gedeck) and Mariellina (Iaia Forte); the first is home sewing and suddenly drops her work to go out and run along a deserted street at night, while the second is in her room,

nostalgically looking at photographs and leafing through the pages of a book. Inside the book she finds a flight ticket; she smiles while tears run over her face.

Both clips were mounted on Tobii T/X Series, which we used with Tobii Studio 2.0.4, with 60 Hz sampling rate and 16.6 ms tracking rate. We took participants through a calibration phase, providing two screenshots with very basic instructions. Calibration was immediately followed by presentation of the selected clips. To an overall 30 participants, we presented alternately Clip 1 followed by Clip 2, or Clip 2 followed by Clip 1. The choice of participants was made according to two, basic criteria: they had to be between 20 and 50 years of age; and they had to have no significant sight correction. We did not require any specific level of education, interest or occupation, and we did not provide any information about the experiment. The instructions given to participants were limited to: 'You will be watching two different clips from the same film. Please, watch these clips without moving your head too much'.

On the whole, we recorded 30 tests, but had to discard six sets of data due to inaccurate fixation. This left us with a final group of 24. We then grouped the recordings for each of the two clips and decided to perform a twofold analysis on the data. After isolating 50, two-second micro-sequences, we created *heatmaps* and *gaze plots* and on these two features we grounded our data analysis.[7] The 50 micro-sequences allowed us to cover the whole 100 seconds for each clip; they also allowed us to base our observations not on still images but on micro-movements, which both heatmaps and gaze plots help to understand.

In the two images below (Figures 1 and 2), the different colours and their increasing/fading intensity indicate the fixation points of the 24 participants: from red to orange, yellow and fading green, areas of interest (AOIs) clearly emerge from this and the other heatmaps we obtained with our tests. On the whole, heatmaps tend to confirm the universally-known fact that viewers focus primarily on the centre of the screen for image perception, although additional fixations and their respective intensity/duration can be assumed as a valuable datum to infer the importance and order of perception of various elements.

A *gaze plot* is like a chart drawn on an image or micro-sequence, made of lines and circles that pattern the gaze control of individuals or groups. Gaze plots offer a joint visualization of fixations (and their respective dwell time) and gaze motion, i.e. the journey that the eyes make over an image or sequence. Fixation points and dwell time are displayed by means of circles, whereas gaze motion is defined by saccades, i.e. the fast movements the eyes make between fixations. Saccades are here identified as the lines connecting the circles, whose sequential order is defined by numbers from one onwards. Figures 3 and 4 below provide clear examples.

In our experiment, if heatmaps were extremely useful to detect AOIs and, therefore, visual priorities over the 50 micro-sequences, gaze plots provided a confirmation of those priorities while also allowing us to monitor gaze control. For the sake of our experiment, the same data visualized in the form of heatmaps and gaze plots were observed by means of charts and graphs, which yielded the data we needed to draft our ET-induced audio description. This stage will be the object of the next section.

Stage two: drafting ET-induced audio descriptions

Creating AD is like walking a tightrope. (Kruger, 2012, p. 3)

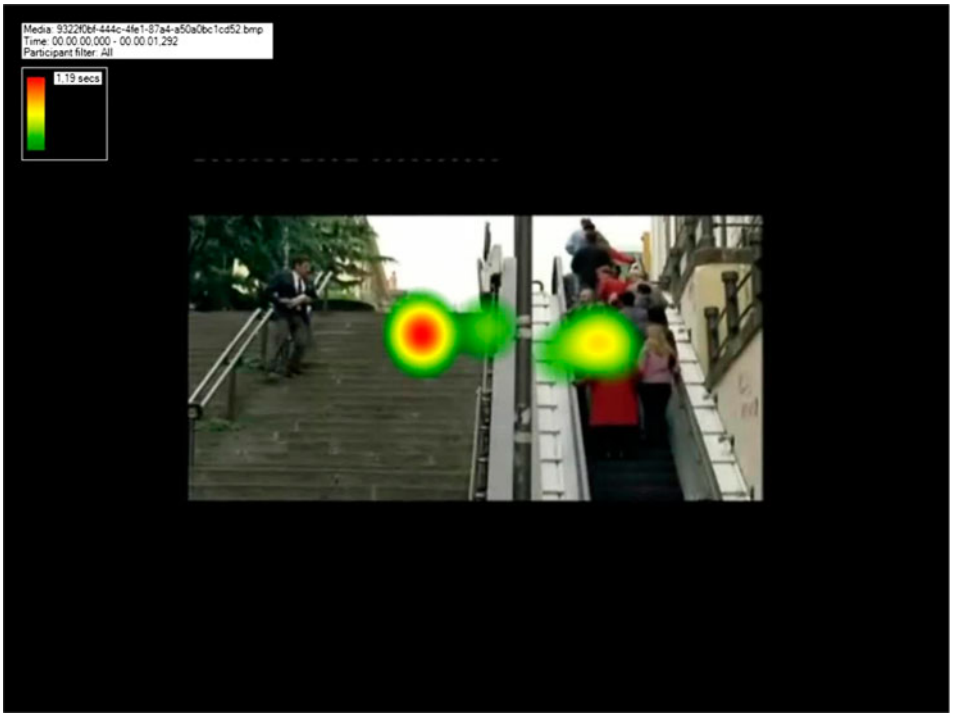

Figure 1. Heatmaps for Clip 1, long shot of chase over stairs.

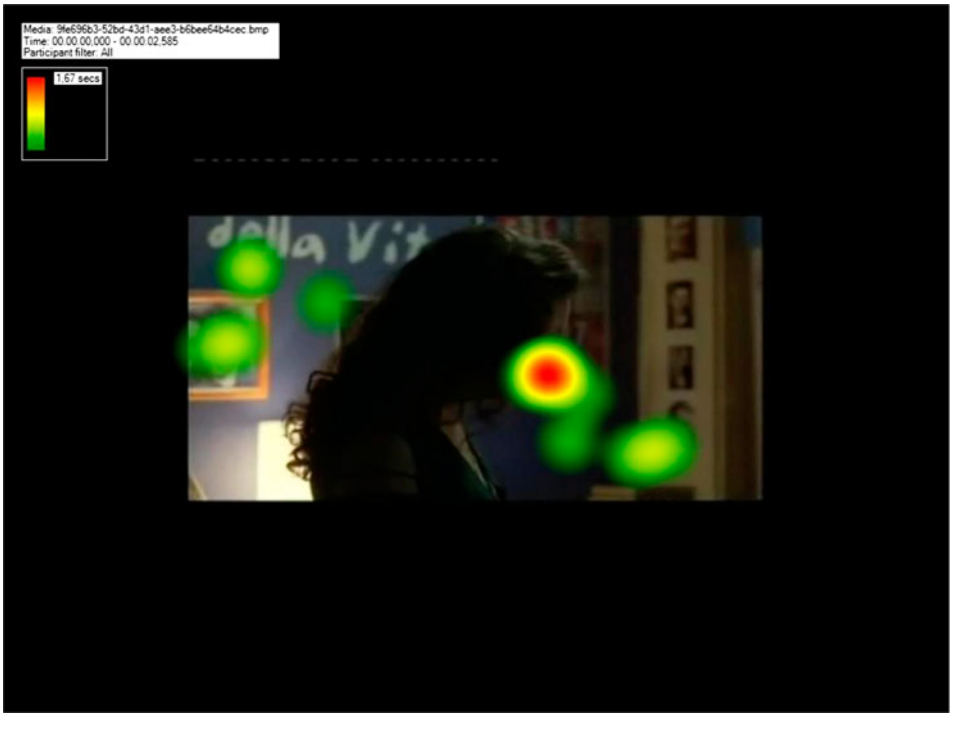

Figure 2. Heatmaps for Clip 2, medium shot of female character.

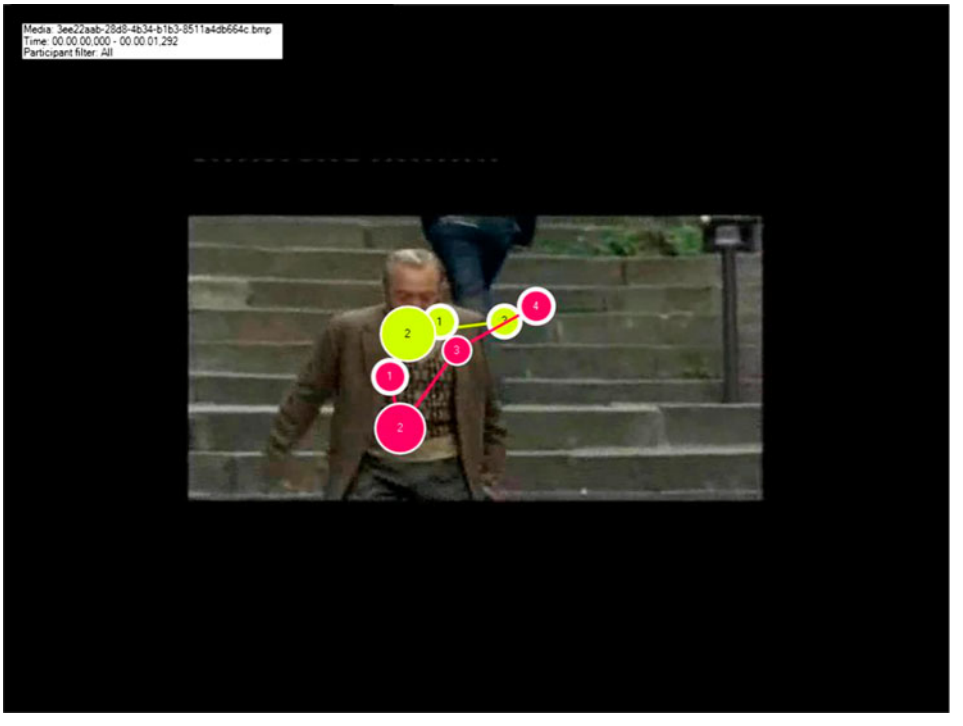

Figure 3. Gaze plot for Clip 1, medium shot of man during chase.

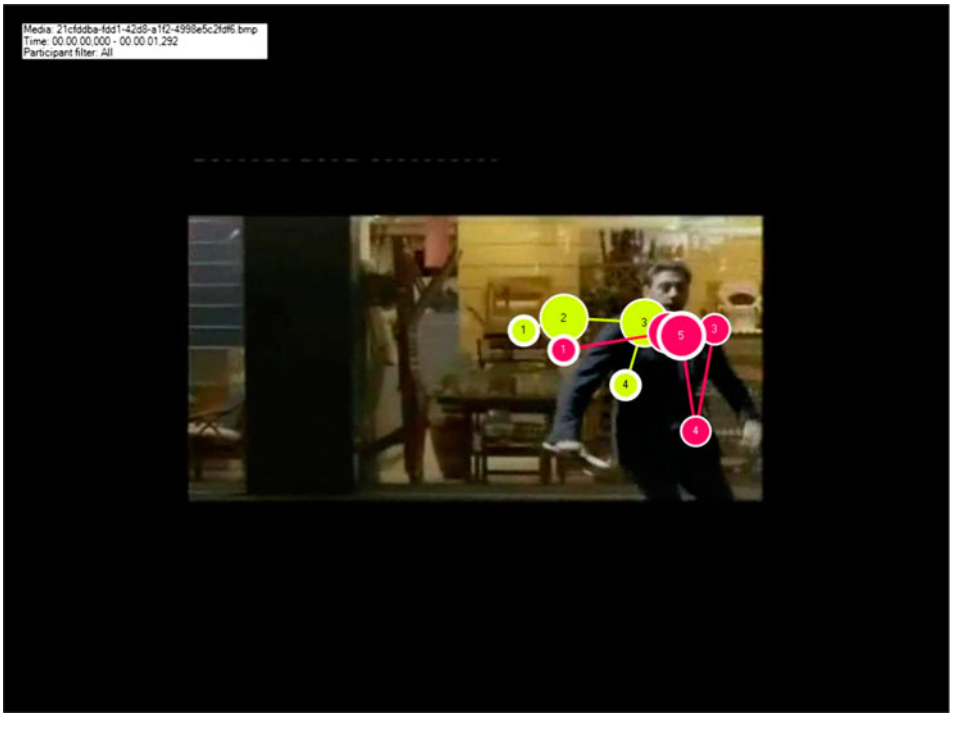

Figure 4. Gaze plot for Clip 1, inside the shopping arcade.

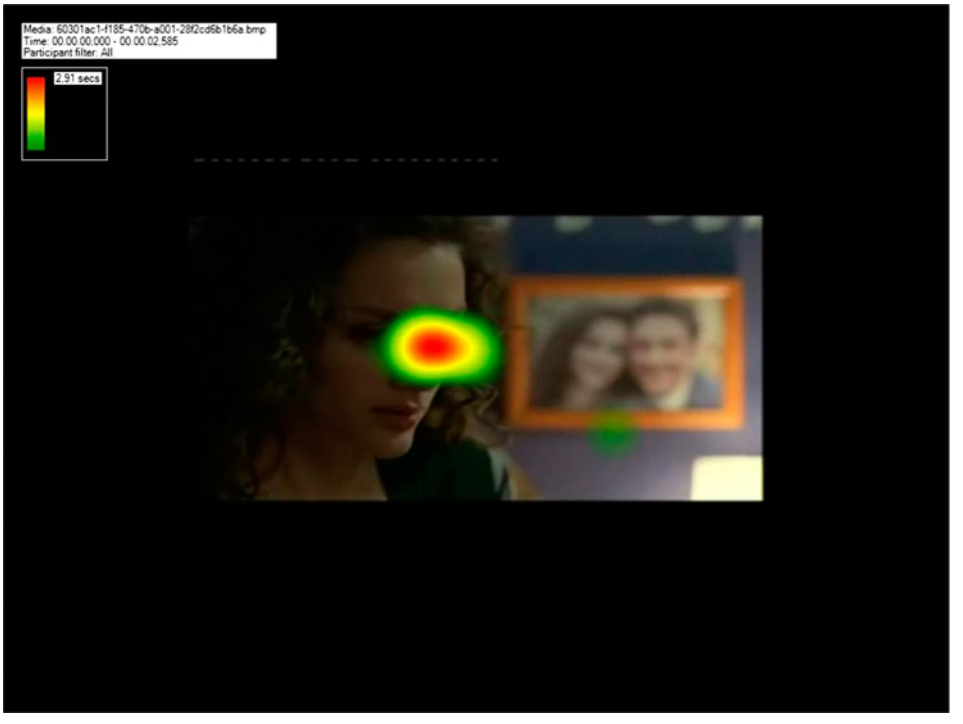

Figure 5. Heatmap for Clip 2, close up of female character.

The next step in our experiment implied 'translating' all the data gathered through the ET tests into strings of words which ought to: reflect the priorities of sighted viewers in terms of selection of visual information, fixation length and gaze control; and appropriately function as descriptions for the blind and visually impaired. If the first point was indeed difficult to achieve in itself, the second was even more insidious, as the shift from the observation of heatmaps and gaze plots to the drafting of 'viable' descriptions was no easy business. Indeed, if compared to the task of audio describers working under normal conditions, i.e. relying solely on their own perception and interpretation of visual stimuli, our task was somehow made easier by the identification of salient visual information through the ET tests. Nonetheless, a number of issues remained open and had to be tackled when drafting our ET-induced ADs:

(1) where to draw a line between information to be necessarily conveyed or to be possibly excluded from our descriptions;
(2) how to combine the selected information without adding any connotation or nuance deriving, for instance, from syntactic structuring; and
(3) how to give a narrative structure to our data while also trying to stay as 'neutral' as possible.

Although these questions can be said to apply loosely to AD writing in general, they deserved maximum attention when drafting ADs whose aim was to stay away from individual interpretation and textual intervention.

For more clarity, let us discuss each of the three issues above in sequence.

The much discussed question of *what* is to be conveyed in AD (see Greening & Rolph, 2007; Snyder, 2005; Vercauteren, 2007) was in our case smoothed out by our recourse to the information provided by the ET tests. However, the equally thorny issue of *how much* to convey remained open: if turning the shades of colours provided by heatmaps and the size/sequence of circles defined by gaze plots into data for AD had not been a problem, making a qualitative selection over the quantity of these data was neither straightforward nor easy. However, since we were to use the ET-induced ADs in comparison/contrast with the already existing, normally produced ADs, we gauged our selection on the latter. In other words, we calculated the amount of time covered by the normally produced AD in the selected clips and replicated that coverage in our new ADs. Moreover, in order to make the two sets of descriptions truly comparable, we had them all recorded by the same voice. This also implied making some additional adjustments to our newly generated AD, to make sure duration and time coverage were perfectly equal.

However, these finishing touches were applied *after* the ET-induced AD had been drafted, a task which was performed bearing clearly in mind points 2 and 3 above. To start drafting our texts, we decided to create a series of minimal narratives (Labov, 2001), featuring all elements that had scored at least 70% of fixations (out of all participants) and following the paths determined by means of gaze plots. This meant including elements which a traditionally-drafted AD would probably exclude, in virtue of their being 'secondary' (peripheral objects) or hard to describe (face expressions and the changes thereof, as exemplified by Figure 5). This also implied including elements – such as those appearing at the centre of the screen and those embodying main movements – that are generally considered 'automatic' points of fixation, therefore not involving any specific cognitive effort.[8] However, since every film, and every scene within a film, has its own rhythm and focal elements, and since they can also be purposefully emphasized by the director, we can say that each element above the 70% level carries a specific visual and narrative meaning that is worth reflecting in the AD.

Moreover, as our new AD was data-driven, we decided to limit the addition of any possible nuance or connotation by keeping the syntactic complexity towards its lowest limits, using one temporal juncture or two (Labov, 2001, p. 63) within each sentence and coordinating clauses only when breaking them up into autonomous sentences proved inconvenient. By so doing, our ET-induced AD resulted in a sequence of brief and clear narrative units.

Once the two ET-induced descriptions for Clip 1 and 2 from *Tris di donne* had been so drafted, we gauged them against the existing AD for equal length, as described above. With minor deletions and adjustments, the four ADs (two for Clip 1 and two for Clip 2) were now analogous in length and ready to be recorded and tested with end users.

Stage three: presenting the two types of AD to blind individuals

Audio description is the art of narrating what a blind person cannot see when enjoying a film. However, audio description is also, without any doubt, the art of sound (Fryer, 2011). One of the main reasons for the still unpractised recourse to text-to-speech technologies in AD lies in the great importance end users place in the quality of voice and sound. Being aware of the great sensitivity of the blind to these issues, and approaching the final stages in our project, we strived to ensure

Table 1. Questions following the viewing of Clip 1.

CLIP 1 – THE CHASE
1a. Would you say this description was clear?
2a. Could you briefly tell us what happens in this sequence?
3a. Was anything missing from the description?
4a. Is there any detail provided in the description that you deem irrelevant?
5a. Do you think you were able to follow the action?

appropriate sound and voice quality when recording our ADs. Above all, we aimed to ensure uniform conditions in the auditory perception of the two types of AD we were to test. To this end, we had all four ADs re-recorded in a studio with the same voice. We then proceeded to set up an appropriate environment for our final test: we made arrangements with the Macerata section of the major Italian association of the blind and visually impaired, UICI (*Unione Italiana dei Ciechi e degli Ipovedenti*). President Mirko Montecchiani was eager to collaborate and offered us the opportunity to make use of a room within the UICI premises in Macerata. We fixed a date and asked Mr. Montecchiani to summon as many blind individuals as possible for that date.

Eight blind individuals enthusiastically accepted to participate in our experiment; all of them were totally blind, most of them from birth. Only two out of eight had developed blindness at a later stage, but had been totally blind for over two-thirds of their lives. Their age ranged between 21 and 62, with four women and four men altogether.

On 16 October 2010, two hours before administering the test, we set up our equipment in the selected room. Our experiment started when each participant came into the room, sat in front of the table equipped with the laptop and wore the professional headset we had prepared. In order to have our participants familiarize themselves with the setting and the experiment, we asked a short series of general questions before they were presented with the AD clips. The questions aimed to elicit information about the participant's age, nature and cause of their blindness, knowledge and experience of audio description. The replies were recorded on a voice recorder.

We then introduced the experiment by providing the following instructions:

> You will now listen to two, 100-second clips from the same film, accompanied by audio descriptions. The film title is *Tris di donne* and the two clips have been purposefully selected to ensure that there is little dialogue and plenty of visual information to describe. At the end of the second clip, we will be asking you some questions about the audio descriptions.

Table 2. Questions following the viewing of Clip 2.

CLIP 2 – THE TWO WOMEN
1b. Would you say this description was clear?
2b. Could you briefly tell us what happens in this sequence?
3b. Was anything missing from the description?
4b. Is there any detail provided in the description that you deem irrelevant?
5b. Do you think face expressions should or should not be described?

Before carrying out the experiment, we divided the participants in two groups, ensuring that age range was largely covered in both groups (regardless of gender) and that one of the two who had become blind later in life was in each group. Subsequently, we organized the tests, making sure that four individuals listened to the ET-induced AD for Clip 1 and the normally drafted AD for Clip 2, while the remaining four listened to the normally drafted AD for Clip 1 and the ET-induced AD for Clip 2. After they had listened to both clips, each of the eight participants was asked questions to evaluate their comprehension and overall reception of the clips through the different ADs.

Stage four: gauging end users' reception of the two ADs

The completion of the second experiment within our project, i.e. the testing of two different types of AD with blind individuals, did not coincide with the end of the experiment itself. It implied transcribing and analysing the recorded questions and answers each participant had provided, before and after listening to the AD clips. This lengthy process yielded surprisingly interesting results, which led us to confirm our initial, apparently oxymoronic hypothesis (see 'The project: background and objectives'), but also to obtain useful information that could possibly be poured into the practice of AD writing. In order to provide a meaningful account of the questionnaires and, above all, to highlight the most significant results they had yielded, we will structure this report along three main issues: the amount of information to be provided in AD; the importance of information sequencing; and the description of emotions through facial expressions.

On the whole, every participant in the project proved to be well aware of the importance of audio description for the appreciation of an audiovisual text and they all stated they would like to have more opportunities to enjoy audio described television and cinema in Italy. Tables 1 and 2 above features the questions each participant was asked after listening to the audio descriptions. Although the questions were repeated for each clip, we listed them twice here, with numbers and the letter 'a' for Clip 1 and 'b' for Clip 2, so as to provide specific comments in the next paragraphs.

How much to describe

As a direct reply to questions 4a and 4b, but also more or less directly when replying to other questions, all eight participants stated that the more details provided, the freer a blind individual feels in forming his/her own reception of a scene or film. Quite interestingly, the most positive reactions to the richness in detail were provided with reference to clips with ET-induced AD. As stated above, the ET-induced ADs deliberately included references to visual information that would probably be deemed marginal in traditional AD writing, but which scored a high number of fixations by our sighted viewers in their ET tests. Whether these fixations were due to the actual salience of these elements or rather to their being merely in the eye trajectory from one element to another, for instance when tracking an action, we decided to add them to the descriptions anyway and see what the reactions of the blind respondents would be. For instance, in Clip 2 one of the two women, seen inside her house in the opening shots, suddenly takes her coat and goes outside. She runs through a deserted street at night and this is undoubtedly the main action she performs. The ET tests,

however, revealed a lot of saccades and a fairly good amount of fixations on the visual elements that appear along the street, cars and trees in particular. For this reason, and having some additional time available, we decided to add these references to the AT-derived AD, whereas no such reference appears in the traditional AD, which is provided below as a reference:

Traditional AD

Josephine esce di casa di gran fretta e corre. E' sera: la via, illuminata dai lampioni, è completamente deserta.
[Josephine gets out of the house in great haste. It is nighttime: the street is lit by streetlamps and totally deserted.]

ET-induced AD

Josephine corre lungo una strada pedonale buia, con macchine ferme sul lato sinistro e una fila di alberi sul lato destro.
[Josephine runs along a dark, pedestrian street, with cars parked on the left hand side and a row of trees on the right.]

With reference to this type of choice, Adriano, aged 62, stated that 'the more details are provided, the greater is the involvement in the action'. Serena, aged 20, made a specific comment to the AD above:

> Perhaps saying that there are cars on the left and trees on the right hand side of the street is not essential for understanding. However, I am not bothered by these elements. If you can add an extra element, why should you not do so? I want to be able to know as much as possible about the director's choices.

Franco, aged 49, in reply to 4a (after listening to the ET-induced AD) said: 'If details are not provided, you have to use your imagination. If you give more details, I can follow the action. Less details and I let my imagination travel, but sometimes I just get the whole sequence wrong'. In commenting on the extra details provided in the ET-induced AD for Clip 1, Bruna, 57, stated: 'You tell me Franco hides behind a newspaper when he gets out of the home appliance shop. This sounds irrelevant but it is through this information that I can understand the chase is still under way'.

With reference to the two normally drafted ADs, most participants (six out of eight) did not point to any irrelevant or redundant piece of information, nor did they declare that something could have been avoided. Two out of eight referred to some basic elements that sounded odd to them, as the description didn't seem to adequately link them up with what came right before them, whereas no such claim was made with reference to the ET-induced ADs. In general, these findings seem to support the idea that AD should not be too selective, and that taking into account viewers' visual attention is indeed relevant for AD. Clearly, more evidence would be needed to translate such a claim into action, i.e. transfer it to AD guidelines, especially since the sample of participants involved in this experiment is evidently limited. Larger scale studies are, therefore, not only recommended but certainly needed.

Information sequencing

Clip 1, which features virtually no dialogue but a sequence of actions in the chase of Franco by Matteo, yielded very different replies to the questions asked to the four

plus four blind respondents. Even if two out of the four people who listened to the normally drafted AD declared that the description was clear, when replying to questions 2a and 3a it clearly emerged that they were unable to reconstruct the series of passages that lead the two protagonists from the concrete stairs to the shopping arcade, the home appliance store, the street, the train station and the train. The omission of essential information in the original AD, the audience of which is not informed about the passage from an outdoor staircase to a shopping arcade and a home appliance store, generated a lot of confusion. While commenting on this (in reply to question 2a), Giulia, aged 23 stated: 'It seems that the scene takes place in the street. Then one of the two men gets into a store and the other is outside, looking at a shop window. Then they meet again, but where? Then one runs after the other and we get lost when they are inside the train. Perhaps there is a scene change at some point?'

One (Giuseppe, 69) of the four respondents who listened to the ET-induced AD remarked that perhaps a few details could have been omitted. He said: 'Telling me there are a lot of people in the train station is not important, I would rather listen to the noises, without any comment'. However, he also added: 'The good thing about this description is I understand where both men go, and it sounds like a quick series of movements. I can follow them.' Giuseppe's comments are precious and point to aspects that are essential for successful AD writing. Covering with descriptions all those sounds which can speak for themselves is generally to be avoided, just like providing too much information (see 'Stage two: drafting ET-induced audio descriptions'). The tendency to over-describe may also be due to using data on attention allocation, which seem to be all relevant, especially in the absence of dialogue. A careful use of these data is obviously necessary, just as it is necessary to consider the precise sequence of the saccades and smooth pursuits. This is precisely what can be inferred by Giuseppe's second remark: his full appreciation of the quick series of movements implies that the actions and movements have been appropriately sequenced.

Having stated earlier that the duration of the two ADs for the same clip was equally balanced, the greater amount of information (and details) stated by the respondents to have been provided in the ET-induced AD may appear suspicious at this stage. How could more information be given in the same amount of time? The reply to this question is straightforward: resorting to minimal narratives and trying to eliminate all possible 'extras' in the data-driven AD has led to the creation of short, clear-cut and semantically denser sentences. Also, and perhaps more importantly, the data provided by the ET tests have pointed to visual priorities that do not always correspond to the information provided in the traditional AD. One may here remark that, had the traditional AD been different, the results of the experiment may have also differed greatly. This is indeed unquestionable but traditional AD, especially in countries/contexts where the very practice and distribution of AD has so far been limited, does not follow rigid canons and guidelines, which could perhaps allow for a classification of the traditional AD used for this experiment in terms of (poor) quality. One may also observe that shorter sentences allow for easier information sequencing and processing, which is equally unquestionable. However, since the main focus of this article is the methodology that was tested with this experiment, we shall not linger on considerations of sentence length and complexity, or bottom-up processes, which nonetheless deserve investigation when it comes to applying ET research to audio description.

To proceed with our analysis, the following is an excerpt from the two ADs for Clip 1, covering the same time span:

Traditional AD

Franco inizia a correre, si volta e poi entra in un negozio di elettrodomestici. L'uomo si ferma a guardarsi intorno di fronte a una vetrina. Franco esce dal negozio e nasconde il viso dietro a un giornale. L'uomo si volta, lo riconosce.
[Franco starts running, he turns around and then enters a home appliance store. The man stops and looks around himself in front of a shop window. Franco leaves the store and hides his face behind a newspaper. The man turns around and recognizes him.]

ET-induced AD

Franco si volta. Correndo, si immette in una galleria di negozi. Entra in un negozio di elettrodomestici e sale la scala mobile. Esce davanti a una vetrina con i cartelli dei saldi. Si copre il volto con il giornale. L'altro, fermo e di spalle, si volta e lo vede.
[Franco turns around. Running, he gets into a shop arcade. He enters a home appliance store and steps onto the escalator. He exits, a shop window with sales posters in front of him. He hides his face with a newspaper. The other man stands still, his back on Franco. He then turns around and sees him.]

Both excerpts above reflect the quick movements of the two characters; the sequence described is some 14 seconds long. The differences in the construction of the descriptions are self-evident and, besides pointing to a different choice of information, they highlight that the data from the ET tests have led to developing a more comprehensive and linear information sequencing, which the eight participants in our experiment had demonstrated through their feedback.

These and the findings described in the following paragraphs seem to suggest that our main initial hypothesis may be confirmed, although this still needs to be established in a more comprehensive study, controlling for variables such as sentence structure and length.

Describing facial expressions

Last but not least, our survey of the most significant issues that emerged from our testing of different ADs with blind individuals focuses on the descriptions of facial expressions and their reception.

In its *De Oratore III*, Cicero says 'Ut imago est animi voltus sic indices oculi'[9], i.e. the face is the image of the soul as the eyes are its interpreters. Indeed, in an excerpt like the second one (Clip 2), which was presented to the eight blind respondents with two different ADs, faces bear the burden of conveying varying emotions. And it is those emotions that make the film narrative develop. This is what emerged, first of all, from the ET tests: at least 90% of the fixations across this excerpt are on the characters' faces, with durations going well beyond those recorded for Clip 1 and saccades often reflecting movements only across the faces.

The figure above shows an example of what has just been said. Moreover, it has to be reported that in this shot the woman is almost in extreme close up, whereas most other shots see her in close up and also, very frequently, in medium shots.

Conveying expressions in AD is no easy task, especially as one of the most commonly shared principles in AD writing is that expressions should be described physically, not by attempting to interpret emotions. Thus, striking a balance between

saying what a face expresses and not providing any arguable interpretation is extremely difficult. The two ADs that we presented to our blind respondents both contained references to the characters' facial expressions. However, as we can see below, these references neither occur at the same time nor are they expressed in analogous ways. The first sentence for the two ADs below occurs at the very beginning of Clip 2. The traditional AD starts off with what sounds like an interpretation of the character's thoughts, which has been inferred from her expression. However, this opening sequence is perhaps too quick to justify such an interpretation, a datum which is confirmed by the ET-tests: the viewers' eyes lay first of all on Josephine's lap, where her hands and needling work are. They then quickly moved to her face and then followed her rapid movement out of the room, all of this in no more than five seconds. In this case, we can say that, differently from other occurrences, interpreting the ET data has led us to avoid providing an over-detailed description of a facial expression, which also seems irrelevant to understand the action at this point.

Traditional AD

Colta da un improvviso pensiero, Josephine posa forbici e abito, si alza di scatto e afferra il soprabito. [. . .] Tra le pagine del libro c'è il coupon di un biglietto aereo. Luisa sorride, poi trova anche un biglietto con su scritto "quando torni mi dici che sto guardando". Si commuove e chiude il libro, tenendolo in grembo fra le mani.
[Seized by a sudden thought, Josephine leaves aside scissors and dress, stands up briskly and grabs her coat. [. . .] Among the book pages she finds a flight ticket. Luisa smiles, then she also finds a note which says "when you're back, tell me what I'm looking at". She is moved and closes the book, keeping it on her lap, in her hands.]

ET-induced AD

Josephine è seduta e ha lo sguardo su ciò che sta cucendo. Lascia il lavoro, si volta ed esce di fretta. [. . .] Nel libro ingiallito trova un biglietto aereo. Lo apre e legge, accennando un sorriso. Trova anche un biglietto con scritto "quando torni mi dici che sto guardando". Il suo sorriso si fa più grande e si commuove. Chiude il libro dalla copertina ingiallita.
[Josephine is seated, her eyes on what she is sewing. She leaves it aside, turns around and quickly goes out. [. . .] Luisa finds a flight ticket inside the yellowed book. She unfolds and reads it. She smiles faintly. She then finds a note with the words "when you're back, you tell me what I'm looking at". Her smile broadens, she is moved. She closes the book whose cover is yellowed.]

The second segment of the two ADs above refers to a rather static sequence: Luisa is sitting on a chair, her face and breast in close up, shot from the right. Attempting to comply with commonly-shared guidelines for AD writing (in Italy and, most of all, abroad), which restrict the recourse to detailed face descriptions, the traditional AD only refers once to Luisa's smile, when she finds the plane ticket inside the old book. However, her smile is not static; it broadens visibly as she reads the note. This datum is reflected in the ET tests, the analysis of which revealed a clear series of saccades from Luisa's face to the objects she is looking at, as well as long and detailed fixations on her eyes and mouth after reading both the ticket and the note. This is what led us to add references to her changing smile in the ET-induced AD, a detail which was appreciated by three out of the four respondents who listened to it.

In reply to question 2b, Giuseppe, 66, stated that he had been particularly struck by Luisa's expressions: 'The tears that come to Luisa and the smile that changes really struck me. Without the description I would have had no idea of her feelings'.

As a general comment to the Clip (reply to question 1a), Bruna, 57, said: 'I think I understood it all. I could dive into this scene, which is good. I especially liked the description of Luisa, her smile, her room, the yellowed book'. Finally, as a reply to question 5a, Serena, 20, stated: 'I think facial descriptions are essential to understand the psychological state of each character, which otherwise we can only guess. They are also useful for our personal growth, or so I think. Descriptions of changing expressions allow us to follow the film and grasp its deep meaning.'

Conclusion

The end of the account of this complex experiment brings us back to the beginning, but with a new perspective: the seemingly oxymoronic nature of the relationship between eye tracking research and audio description has been dismissed and, although further studies are clearly needed, ET research and the evaluation of attention allocation by sighted viewers have proven extremely relevant and useful for the study and practice of AD. This is certainly one of the main results that can be drawn from this experiment; the eye tracking data were measured and analysed, then applied to the creation of ADs that were finally tested with the blind in contrast with normally drafted ADs. The wide consensus expressed by the participants, as well as an overall better understanding of ET-induced over normally produced ADs – as reported in the oral questionnaires – can be taken as valuable proof of the relevance of ET research for audio description. However, this experiment ought to be taken as a sort of pilot study, given the small number of blind participants involved and being the first of its kind to be reported. Indeed, more research is needed, making sure that: a wider sample of the blind and visually impaired population is taken into account; different film genres, scenes and forms of interaction between dialogue/sound/silence are analysed; and different linguistic strategies are tested to give textual substance to ET data. Further research should also be carried out to investigate the socio-cultural specificity versus universality of the results obtained by this and further studies, for instance by replicating experiments across different countries/cultures.

On a more general level, we can say that one of the major strengths of the experiment here reported lies in its revolving around, and relying upon, the input provided by users from beginning to end. Indeed, it has considered the actions and reactions of viewing as well as blind individuals, testing their response to visual and verbal stimuli in their cinematic experience. Besides the social relevance of user-centred research of this kind, we should not overlook the potential impact of any improvement that can result from it. In line with this, and with the need to locate research within more or less clear-cut domains, we could say that the complexity of the experiment here discussed makes it relevant for audiovisual translation studies, accessibility, eye tracking and cognitive studies, and perhaps a few more domains. Its starting and ending with people's responses and its having an impact on the improvement of processes also makes it a good instance of action research. As McNiff and Whitehead (2009, p. 11) define it, 'action research lends a new dimension, because it is about processes of improvement, and making claims that something has improved'. And something has already improved after this project was carried out: the results obtained through the reception tests with the blind have been poured into the revision of AD guidelines for an international company providing access services.[10] In particular, instructions on information sequencing, on descrip-

tions of facial expressions and sentence structure in AD have been added and integrated, with examples from the reception tests carried out with the blind. These and other results have also been shared with the audio description department of the major Italian broadcaster RAI, in view of the restructuring of their audio description service.

Notes

1. Among the studies published so far, see the pioneering work of Pilar Orero (Orero & Vilarò, 2012), only published for the first time in 2012 but carried out since 2008. See also, among others, Krejtz et al. (2012) and Kruger (2012).
2. Research on audio description for the 'Digital Television for All' project was carried out mainly in Catalonia, Poland and Italy. The investigators involved were: Pilar Orero and Anna Vilaró from Universitat Autònoma de Barcelona, Catalonia; Agnieszka Chmiel and Iwona Mazur from Adam Mickiewicz University in Poznan, Poland; and Elena Di Giovanni and Sara Fusari from Università di Macerata, Italy.
3. The team included the author of this paper, in charge of the Italian unit of AD researchers for the 'Digital Television for All' project, and Sara Fusari, a postgraduate student at the University of Macerata from 2008 to 2011.
4. A clear and simple definition of iconic signs with reference to the audiovisual media, although here referred to TV in particular, is provided by Jonathan Bingell: 'Many of the television's visual signs resemble the people, things and places which they represent in both fictional and non-fictional programmes. Signs which resemble their object in this way are called iconic signs, to distinguish them from signs which themselves have no necessary relationship to what they signify. The word 'cat', for example, is a symbolic sign, meaning that the letters on the page or the sound of the word 'cat' is arbitrarily used in English to signify a particular type of furry four-legged animal. A television image of a cat, however, closely resembles the real cat which it represents, and is thus an iconic sign' (Bingell, 2004, p. 87).
5. See, for instance, Benecke (2007). Audio Description: Phenomena of Information Sequencing.
6. See, for instance, the *Marie Antoinette* Project mentioned earlier.
7. A *heatmap* is the area of focus for an image (or micro-sequence); it is displayed with different shades of colour, which identify the main focus of fixation for one or more individuals.
8. We are aware that bottom-up processes in visual attention influence the viewers' behaviour.
9. *De oratore* is a three-volume dialogue written by Cicero in 55 BC. The critical edition used for and cited in this paper was published in 1995, by Kazimierz Kumaniecki.
10. SubTi Access is an international company based in Italy and providing accessibility to film festivals and TV networks.

Notes on contributor

Elena Di Giovanni is Lecturer in Translation at the University of Macerata (Italy), where she is also Director of the Language Centre. She holds a degree in specialized translation from the University of Bologna at Forlì and a PhD in English and audiovisual translation from the University of Naples Federico II. She has taught audiovisual translation theory and practice within several MA programmes: University of Bologna/Forlì, University of Parma, IULM (Milan), Universitat Autònoma de Barcelona (Spain) and Roehampton University, London (UK). She is also Director of the international MA programme in Accessibility to Media, Arts and Culture (University of Macerata).

Her research interests include translation as intercultural communication, translation and postcolonialism and audiovisual translation in all its forms. She has published extensively on

subtitling, dubbing, audio description and other forms of audiovisual translation. She has been working as a professional audiovisual translator for over 20 years.

References

Benecke, B. (2007). Audio description: Phenomena of information sequencing. In MuTra (Ed.). *2007 – LSP Translation Scenarios: Conference Proceedings*. Retrieved January 18, 2013, from http://www.euroconferences.info/proceedings/2007_Proceedings/2007_Benecke_Bernd.pdf.

Bingell, J. (2004). *An introduction to television studies*. New York: Routledge.

Cicero, M.C. (1995). *De oratore (Fasc. 3)* (K. Kumaniecki Ed.). Stuttgart: Stereotypa Ed.

Di Giovanni, E. (2009). Translation, cultures and the media, special issue of *EJES. European Journal of English Studies*, *12*(2), 121–225.

Di Giovanni, E., Orero, P., & Agost, R. (2012). Introduction to multidisciplinarity in audiovisual translation. *MonTI, 4*(2012), 9–49.

Fryer, L. 2011. Audio description: the art of sound? Paper presented at the 3rd Advanced Research Seminar on Audio Description *ARSAD*. Universitat Autónoma de Barcelona, Spain (24–25 March).

Greening, J., & Rolph, D. (2007). Accessibility: Raising awareness of audio description in the UK. In J. Diaz Cintas, P. Orero & A. Remael (Eds.), *Media for all. Subtitling for the deaf, audio description, and sign language* (pp. 127–138). Amsterdam: Rodopi.

Krejtz, I., Szarkowska, A., Krejtz, K., Walczak, A., & Duchowski, A. 2012. Audio description as an aural guide of children's visual attention: evidence from an eye-tracking study. In *ETRA '12, Proceedings of the Symposium on Eye Tracking Research and Applications* (pp. 99–106). New York: ACM.

Kruger, J.L. (2012). Making meaning in AVT: Eye tracking and viewer construction of narrative. *Perspectives Studies in Translatology*, *20*(1), 67–86.

Henderson, J.M., Brockmole, J.R., Castelhano, M.S., & Mack, M. (2007). Visual saliency does not account for eye movements during visual search. In R.P.G. van Gompel et al. (Eds.), *Eye movements: A window on mind and brain* (pp. 537–562). Oxford: Elsevier Ltd.

Labov, W. (2001). Uncovering the event structure of narrative. In *Georgetown University round table on languages and linguistics (GURT) 2001*. Retrieved January 31, 2013, from http://www.ling.upenn.edu/%7Ewlabov/uesn.pdf.

McNiff, J., & Whitehead, J. (2009). *Doing and writing action research*. Los Angeles: SAGE.

Millner, D., & Goodale, M. (2006). *The visual brain in action*. Oxford: Oxford University Press.

Monaco, J. (2000). *How to read a film. Movies, media, multimedia* (3rd ed). New York: Oxford University Press.

Orero, P., & Vilaró, A. (2012). Eye tracking analysis of minor details in films for audio description. *MonTI*, *4*(2012), 295–319.

Poli, A. (2009). *Cinema e disabilità visive. L'esperienza filmica senza colore*. Milano: FrancoAngeli.

Rainer, K., Smith, T.J., Malcolm, G.L., & Henderson, J.M. (Eds.) (2009). Eye movements and visual encoding during scene perception. *Psychological science*, 20(1), 6–10.

Snyder, J. (2005). Audio description. The visual made verbal across arts disciplines - Across the globe. *Translating today*, *4*, 15–17.

Vercauteren, G. (2007). Towards a European guideline for audio description. In J. Díaz Cintas, P. Orero & A. Remael (Eds.), *Media for all: Subtitling for the deaf, audio description and sign language* (pp. 139–149). Amsterdam: Rodopi.