Harnessing the Potential of Eye-Tracking for Media Accessibility

Agnieszka Szarkowska (University of Warsaw)
Izabela Krejtz, Krzysztof Krejtz (University of Social Sciences and Humanities)
Andrew Duchowski (Clemson University)

Introduction

This article presents an overview of eye-tracking research studies in the area of media accessibility carried out by the Audiovisual Translation Lab at the Institute of Applied Linguistics, University of Warsaw, in cooperation with the Interdisciplinary Center for Applied Cognitive Studies at the University of Social Sciences and Humanities in Warsaw (SWPS). Media accessibility is here understood as various tools providing access to audiovisual media content for people with sensorial disabilities: in the form of subtitles for those who are deaf or hard of hearing and in the form of audio description (AD) for those who are blind or partially sighted.

The article begins with a brief discussion of various applications of eye-tracking to audiovisual translation and media accessibility, and then continues with a presentation of the following four studies carried out by the authors of this article: (1) on subtitling for the deaf and hard of hearing (SDH) as the Polish contribution to the DTV4ALL project, (2) an on-going study on SDH on digital television, (3) audio description in education, and (4) audio description in art.

1. Eye-tracking, reading and audiovisual translation

Over the last thirty years eye-tracking methodology has been successfully applied to a host of research domains (for an overview see A. Duchowski 2002, K. Rayner 1998). A rapidly growing number of studies is using eye movement behaviour to investigate individual differences in cognitive processes, most frequently in visual attention. A. Duchowski (2002) classifies the role of eye-trackers in research as either diagnostic or interactive. In the diagnostic approach we gather objective numerical data of subjects' attentional processes, whereas in the interactive studies eye movements are used as a tool allowing users to operate many applications such as eye-typing systems for disabled people who cannot type or use a mouse (e.g. P. Majaranta, K.-J. Raiha 2002). Although interactive gaze-contingent subtitles, appearing when our eyes look for them, are just a near future, in our projects we concentrated on the analysis of subtitle reading patterns across different groups of film viewers as well as on the influence of audio description on the film and artworks viewing process.

Eye-tracking has been extensively used in reading studies (see K. Rayner 1998, K. Rayner et al. 2012), particularly for the analysis of monosemiotic texts. It has been found that when reading printed text in English the eyes make quick movements (saccades) from one word to another, lasting for about 30-50 milliseconds (ms) or 7-9 letter spaces, and they stop on words, resulting in fixations, usually lasting for about 100-500 ms (K. Rayner 1998, K. Rayner et al. 2012). The duration of a fixation depends on a number of factors, such as the length of a word (longer words tend to be fixated with higher probability than shorter words), its frequency (the more frequent the word, the less likely it is to be fixated) and grammatical category (content words are fixated 85% of the time, while function words only 35%, see P.A. Carpenter, M.A. Just 1983, K. Rayner, S.A. Duffy 1988, K. Rayner 1998, K. Rayner et al. 2012).

Reading subtitles is a much more complex process than reading monosemiotic texts like a printed page with no illustrations, because along with reading the subtitles, viewers are also watching the moving images accompanied by sound. As a dynamic form of communication, film demands more cognitive resources than watching a scene and reading separately. Watching a film with subtitles combines two types of information processing with different time parameters for fixations and saccades, generally longer fixations and saccades for scenes perception and shorter for reading subtitles. The viewer is scanning a dynamically changing scene, and simultaneously s/ he is tracking the text at the bottom of the screen. This is an example of a dual task causing high switching costs between scene viewing and reading subtitles for some types of viewers.

The presence or absence of sound also seems to be an important factor influencing the way we read subtitles and watch audio described films. When reading subtitles, people usually can hear the film soundtrack, which probably makes the reading different than reading printed text. It has been found that 'mean fixation durations are longer when reading aloud or while listening to a voice reading the same text than in silent reading' (A. Duchowski 2002: 2), which in fact is the case with subtitling. As an extra audio track, audio description also seems to change film viewing patterns. A recent study by A. Vilaró et al. (2012) showed that viewing patterns can be modified by changing the audio track. This result is also corroborated by our findings, reported in I. Krejtz et al. (2012).

2. Previous eye-tracking studies on subtitling

Probably the most famous and widely quoted researcher involved in eye-tracking and subtitling is G. d'Ydewalle. Together with his associates, G. d'Ydewalle examined a number of issues related to the process of reading subtitles in films, such as the automaticity of reading subtitles (G. d'Ydewalle et al. 1991), the importance of presence/ absence of sound when reading subtitles (G. d'Ydewalle et al. 1987, G. d'Ydewalle, I. Gielen 1992), the influence of the number of lines on the reading process (G. d'Ydewalle 1991, G. d'Ydewalle, I. Gielen 1992, G. d'Ydewalle, W. de Bruycker 2007), the familiarity with subtitling as a factor impacting reading (d'Ydewalle et. al. 1991), the language of film dialogue vs. the language of subtitles (G. d'Ydewalle, W. de Bruycker 2007), and the reading of adults vs. children (G. d'Ydewalle, W. de Bruycker 2007). The reading of subtitles has been found to be a largely automatic process: 'reading subtitles is more or less obligatory' (G. d'Ydewalle et. al. 1991: 650), regardless of the presence/ absence of sound and the (in)familiarity with subtitling. According to their findings, people who are familiar with subtitles read them fluently and effortlessly (G. d'Ydewalle, I. Gielen 1992, G. d'Ydewalle, W. de Bruycker 2007).

The eye-tracking measures examined by d'Ydewalle and his associates included the number of fixations per subtitle, fixation duration, saccade amplitude, regressions in a subtitle, subtitle skipping as well as shifts between subtitles and the video picture (see G. d'Ydewalle, W. de Bruycker 2007). The results showed that when reading film subtitles, as opposed to print, viewers tend to make more regressions than in regular reading. The fixation durations in subtitles are also shorter. As for the number of lines, the time spent in a subtitle was proportionally higher in two-line subtitles, which also contain more verbal information difficult to be inferred from the screen, than in one-liners. This prompted G. d'Ydewalle, W. de Bruycker (2007) to state that a more regular reading pattern can be found in two-liners.

G. d'Ydewalle et al. (1987) are widely quoted as having provided evidence for the famous 'six-seconds rule', according to which a two-line subtitle should appear on screen for six seconds – not less because viewers will not be able to read it, and not more because they will re-read a longer subtitle. G. d'Ydewalle et al. (1987: 321) stated: 'For subtitles with two lines the six-seconds rule appears to be excellent for our sample of subjects. In other cases, the existing 'rule of thumb' (the six-seconds rule) of most TV stations is supported by our data'. On a more critical note, it needs to be stressed that the quality of subtitles tested in the studies mentioned above was not always consistent with current subtitling standards. For instance, in G. d'Ydewalle et al. (1987), subtitles consisted of lines up to 32 characters, which is much less than the usual number now (up to 40). In the case of one-line subtitles, they were sometimes displayed in the upper line, and not, as it is usually done in professional subtitling, in the lower line (G. d'Ydewalle, W. de Bruycker 2007: 198) This may have had influence the eye-tracking results. Another major limitation of the abovementioned studies lies in the small number of participants they were based on[19] and thus in their generalizability.

A number of studies on closed captions, in Europe known as subtitles for the deaf and hard of hearing (SDH), was conducted in the US by C. Jensema and his colleagues. In C. Jensema (1998), the participants of the study (578 of whom 205 were deaf, 110 hard of hearing and 262 hearing) were shown video segments subtitled at different speeds. The most comfortable reading speed was found to be 145 wpm, but viewers did not find it difficult to adapt to subtitles of up to 170 wpm. As reported by C. Jensema (1998), frequent subtitle readers – in this case deaf and hard of hearing viewers – were more comfortable with faster subtitles. Hearing viewers, unaccustomed to watching subtitles, preferred slightly slower reading speeds. This stands in contrast with professional SDH standards in Europe, where higher subtitling rates are produced for hearing viewers and lower rates for the hearing impaired.

In another study, C. Jensema et al. (2000a) examined eye movements of six subjects when watching video clips with and without subtitles. In their seminal article they famously stated that 'the addition of captions to a video resulted in major changes in eye movement patterns, with the viewing process becoming primarily a reading process' (2000a: 275). Another major finding was that the higher the reading speed, the more time will be spent on the subtitle area (vs. the image). This result was further corroborated in the study by C. Jensema et al. (2000b), where the viewers looked at subtitles 84% of the time and only 14% at the video picture (and 2% off the video). This, of course, depends on the type of film and the amount of dialogue it contains – it is enough to compare a Woody Allen film with a typical action movie.

On the European ground, Z. de Linde and N. Kay (1999) tested 10 deaf and 10 hearing subjects, recording their eye movements when watching a number of TV programmes such as news, chat show, soap opera, comedy and documentary with intralingual English subtitles. They measured how much time the subjects took to read each word, the number and duration of deflections (jumps between the subtitle and the image), the number of regressions (i.e. re-reading of words

19 Curiously, d'Ydewalle et al. 1987 do not give the data on the number of participants. G. d'Ydewalle, W. de Bruycker (2007) tested 12 adults and 8 children; G. d'Ydewalle et al. (1991) tested 8 American and 18 Dutch students.

and characters) and re-readings of entire subtitles. The authors found a correlation between subtitles and the programme type on the grounds of 'associated discourse styles' (1999: 75), for instance chat shows, as opposed to documentaries, have higher subtitling rates, which in turn has a direct influence on reading speed, which is higher. According to Z. de Linde and N. Kay (1999: 76), 'faster subtitle rates induce faster reading speeds' and 'slow subtitles lead to more re-reading'. Among other factors influencing the reading process were the visibility of the speaker and film editing techniques, such as shot changes, which were the cause of increased re-reading of subtitles. In conclusion, the authors suggest that slow subtitles 'may be more ideal for active location shots' and faster subtitles for 'studio presentations with a static newsreader' (Z. de Linde, N. Kay 1999: 76).

E. Perego et al. (2010) analysed the cognitive processing of a subtitled film fragment based on eye-tracking data combined with word recognition and visual scene recognition. In their study, poor line segmentation in two-line subtitles did not seem to have any negative influence on subtitle comprehension. This finding goes against the results obtained by D.J. Rajendran et at. (2011), who examined the effects of text chunking on subtitling and found that 'chunking improves the viewing experience by reducing the amount of time spent reading subtitles' (2011: 1). The difference in the results may stem from the fact that E. Perego et al. (2012) examined block subtitles whereas D.J. Rajendran et a. (2011) evaluated chunking in live subtitling.

E. Ghia (2012) carried out an eye-tracking experiment testing the role of literal and non-literal translation strategies in interlingual subtitling (foreign film subtitled into subjects' mother tongue). The goal of the study was 'to help disclose the processes of cognitive mapping involved in the perception of subtitled input, and to investigate whether discrepancy of information between ST and TT is liable to affect viewers' overall watching behaviour' (2012: 164). It was found that in the case of non-literal translation strategies, the number of deflections between the subtitle and the image was higher than in the case of literal strategies. This, the author argues, may point to a divergence between the viewers' expectations of a more literal translation and the text they see in the non-literal subtitles.

S. Moran (2012) tested how word frequency and cohesion influence the cognitive effort expended when reading subtitles. Having examined different versions of the same clip (low vs. high frequency and low vs. high cohesion variables), the author reported that subjects watching high-frequency subtitles had significantly lower fixation durations on subtitles, spent more time on watching the video image and had better results in the post-eye-tracking questionnaire. In contrast, participants watching subtitles with low-frequency words and with the low-cohesion variable had to expend more processing effort to read the subtitles, as demonstrated by higher fixation durations. The results are particularly interesting given that the subtitles with low-frequency words were shorter than those with high frequency words. What is more, as pointed out by S. Moran (2012: 216), the results also 'weaken the authority of the currently practiced 6-second rule which assumes a connection between character count and reading speed/ difficulty'. S. Moran (2012: 185) therefore concludes that 'a practical solution to making subtitles more readable is not in reducing the amount of text, but increasing the frequency of lexical items and their combination'.

C. Caffrey (2012) examined the effects of experimental subtitling in the form of pop-up gloss on the reception of TV anime. Pop-up gloss is a note appearing 'on screen to explain or comment on culturally marked items appearing in each of the semiotic channels' (p. 226); it usually explains culturally marked elements from Japanese culture. The analysis of four measures: (1) the percentage of skipped subtitles, (2) the percentage of gaze time in subtitles, (3) the mean fixation duration, and (4) the word fixation probability demonstrated that the presence of pop-up glosses increased the percentage of subtitle skipping, which indicated that viewers' attention was directed from reading the subtitles to looking at the gloss. What is more, participants watching the clips with both pop-up glosses and subtitles felt the subtitles were displayed at higher reading speeds than those participants who watched the clips with subtitles only – in spite of the fact that the subtitle reading speeds were the same in the two conditions. On the positive side, pop-up glosses were helpful in fostering the understanding of culture-specific elements.

To recap, previous eye-tracking studies addressed a number of interesting issues related to subtitling in general and to SDH in particular. The findings provided important insights into our knowledge on how people read subtitles. The vast majority of studies, however, was conducted on relatively small samples of subjects and focused mainly on the English language. Therefore, what is needed now is replicating some of the abovementioned studies as well as addressing new research questions. In our research on SDH reported in this article we have decided to focus on the Polish language by testing Polish hearing, hard of hearing and deaf subjects. We also examine a few recurrent themes in subtitling which have still not been resolved once and for all, such as the edited/ verbatim subtitles debate and the subtitling reading speed.

3. Subtitling for the deaf and hard of hearing – the Polish contribution to the DTV4ALL project

The goal of this study was to establish European standards in subtitling for deaf and hard of hearing, mostly by verifying participants' preferences with their cognitive system capabilities measured through eye-tracking. To address this goal, an eye-movement study was designed and conducted in Poland, Spain, England, Italy and partially in Greece and Germany. The detailed findings of the study are reported in Romero-Fresco (ed., forthcoming). Below we present an overview of our methodology and major findings.

3.1. Rationale

When it comes to subtitling for the deaf and hard of hearing, each country cultivates its own tradition. For instance, on some Spanish TV channels sound information is placed in the top right-hand corner, while in Poland it is traditionally capitalised and situated at the bottom of the screen. Some countries, such as the US, have a long-standing tradition of verbatim subtitles, while others, Poland among them, have come a long way from extensive editing to more faithful and literal subtitles (see A. Szarkowska 2010). The DTV4ALL project addressed the different ways in which SDH is now provided in various countries with a view to creating common European SDH standards based on a number of measures, such as viewers' preferences and eye-tracking data.

3.2. The study material

The study consisted of four major components:

1. a pre-test questionnaire on SDH, where personal data on each participant was gathered;
2. film material of 23 clips, divided into nine parameters (e.g. character identification), each comprising three or two variables (e.g. colours , speaker-dependent placement and tags);
3. three comprehension questions after each clip; the first one about general understanding, the second about the textual content of subtitles and the third about visual elements in the clip;
4. a post-eye-tracking survey on participants' preferences concerning particular subtitling solutions.

The following subtitling parameters, comprising the existing SDH solutions and the ones which could potentially be implemented in the future, were tested in the study:

1. Character ID (colours, speaker-dependent placement, name tags)
2. Subtitling style (verbatim, standard, edited)
3. Position (top, bottom, mixed)
4. Emotions (verbal description, smileys, no description)
5. Sounds (verbal description, icons, no description)
6. Justification (centre, left-aligned)
7. Borders (borders around subtitle letters, no borders)
8. Boxes (black box around subtitles, no black box)
9. Shadows (shadows around subtitle letters, no shadows)

The clips, each lasting about one minute, were taken from the *Shrek* trilogy, dubbed into Polish, so the subtitles tested were in fact intralingual (i.e. based on the translated Polish dialogue).

Thanks to the different ways of collecting data (preference questionnaires, eye-tracking data, comprehension questions) we hoped to obtain information which parameters were optimal for SDH and which could be recommended for further implementation on digital television. Since in this article our major focus is on eye-tracking, below we report data on the eye-tracking tests.

3.3. Participants

The total number of participants in the eye-tracking tests in Poland was 40. They were divided into three groups based on their degree of hearing loss: (1) the deaf, (2) the hard of hearing and (3) the hearing (the control group).

Age group	Deaf	Hard of hearing	Hearing
15-24	1	5	3
25-39	7	9	4
40-59	1	2	2
60+	0	5	1
Total	9	21	10

Table 1. Participants by age and hearing loss

There was a comparable number of male and female participants (46% vs. 54% respectively).

3.4. Procedure

Eye movements were recorded with an EyeLink CL eye-tracking system (SR Research Ltd.). The EyeLink system uses infrared, video-based technology to compute the distance between the pupil centre and corneal reflection. Signals were generated by the eye-tracker at a frequency rate of 500 Hz when a stable corneal reflection was obtainable from a participant. Stimuli were presented on a 17-inch LCD colour monitor with 60 Hz refresh rate. Participants were seated in a chair in front of the monitor positioned at an eye level at a viewing distance of approximately 60 cm, maintained by a forehead and chin-rest.

Participants were presented with 23 clips (each lasting about 1 minute) during which their eye movements were recorded. After three or two clips (i.e. one parameter), participants answered additional three paper and pencil questions concerning their subtitle preferences. Then it was followed by recalibration. Drift corrections were performed before each trial.

Eye-tracking data were analysed for fixations and saccades using the Eyelink DataViewer (SR Research Ltd). For each trial, areas of interest (AOI) were drawn based on prior hypotheses. We compared mean reading time and time to first fixation to the AOIs. For the behavioural data, the percentage of correct answers to comprehension questions was calculated.

Time to first fixation was calculated in milliseconds (ms) and defined as the time which elapsed between subtitle onset and the first fixation that entered the AOI with subtitles. Thus, time to first fixation reflected how much time after the subtitle onset it took the participants to look at the subtitle AOI.

Mean reading time was defined as the period of time that participants spent on reading subtitles compared to the time spent on watching the whole clip. We calculated mean dwell time, i.e. the sum of durations of all the fixations, in the subtitle AOI and compared it to the dwell time of all fixations across the whole clip.

3.5. Results

The study brought about a number of interesting findings. To begin with, we observed differences in reading patterns between deaf, hard of hearing and hearing participants. For instance, hearing people tended to move their eyes from the screen to the new subtitle when a new subtitle appeared, whereas deaf people spent more time in the subtitle area, as if waiting for new subtitles to appear.

According to C. Jensema et al. (2000a: 284), ‘there appears to be a general tendency to start by looking at the middle of the screen and then moving the gaze to the beginning of a caption within a fraction of a second. Viewers read the caption and then glance at the video action after they finish reading'. This regular pattern, the authors argue, is repeated when another subtitle caption appears. In our study, it was more common in the case of deaf participants (see Fig. 1).

Fig. 1: Gaze plot of the reading pattern of a deaf participant when reading verbatim subtitles

Generally, however, the process of reading subtitles in our study was not as linear and regular as C. Jensema et al. (2000a) suggest. When reading subtitles, all the participants made quite a few jumps (called *deflections*, after Z. de Linde and N. Kay 1999) between the image and the subtitle area. On average, they switched more than six times between the image and the subtitle during its presentation, particularly in the case of two-line subtitles (see Fig. 2).

Fig. 2: Gaze plot of the reading pattern of a hard of hearing participant when reading verbatim subtitles

Interesting differences were observed between participants when reading three different types of subtitles: edited, standard and verbatim (Table 2). In our study, verbatim subtitles included every single word from the dialogue, even the words which are usually omitted in subtitling, such as elements of spoken language like vocatives, repetitions, false starts. In contrast, standard subtitles contained most information from the dialogue, but without the elements of oral discourse. Edited subtitles were the ones whose text was simplified and reduced: more complex syntactic structures were shortened, difficult and infrequent words were substituted with their simpler synonyms, while elements of spoken language were removed. Verbatim subtitles had shortest display times and higher reading speed, while edited subtitles were displayed at lower reading speed.

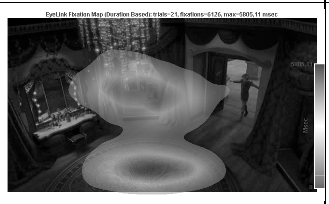
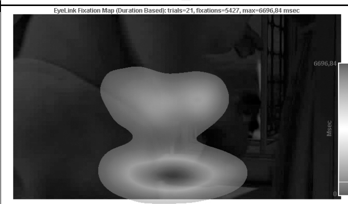
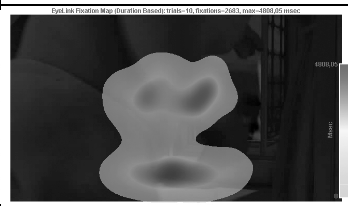
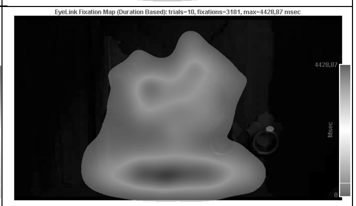
	Edited	Standard	Verbatim
Deaf			
Hard of hearing			
Hearing			

Table 2. Time spent viewing different types of subtitles

Two regular patterns can be discerned when looking at the heat maps in Table 2. First, in comparison with verbatim subtitles, edited captions induced fewer fixations on the subtitles area and more on the video image than verbatim subtitles. This means that edited subtitles allowed viewers to look more at the video picture, whereas with verbatim subtitles people spent more time in the subtitle area, thus having less time to look at the rest of the screen. Second, there is a marked difference between the three groups of participants: hearing people look more at the image and less at the subtitles, as opposed to hearing-impaired people, who spend more time in the subtitle area. In our study, hearing participants made significantly fewer fixations per subtitle (6.58) than deaf participants (8.72). This may be taken as an indication that hearing people are faster readers. It also points to the fact that reading patterns are largely affected by the subtitle reading speed.

Contrary to our initial hypotheses, whereby we expected edited captions to render highest comprehension scores, it was with verbatim subtitles that participants achieved best comprehension results. This, we acknowledge, may be attributable not only to the caption type, but also to the difficulty of the clips and particular questions, so we are far from generalising this result and we believe further tests are needed to examine this area more closely. More information on the results of this study can be found in A. Szarkowska et al. (2011).

The DTV4ALL also tested solutions that are not in common use in the subtitling profession, such as the use of smileys to denote emotions (see Fig. 3) and the use of icons to describe sounds (see Fig. 4), with a view to finding out whether target viewers would like to see them implemented in the future.

Fig. 3: A sample subtitle with a smiley denoting sadness

Fig. 4: An icon of a barking dog used to describe the sound

In the preferences questionnaire, none of the participants selected the smileys as their preferred option of describing emotions and only one hard of hearing viewer opted for icons as the best option of describing sounds. Eye-tracking tests showed that it took people a long time to spot the icon and that many participants did not notice the dog at all. This, of course, may be attributed to the fact that the viewers are not accustomed to such solution of describing sounds. Some other people stated they thought the dog was part of the film. Other reasons may include the quality of the picture and its placement in the top right-hand corner instead of next to subtitles. To sum up, most participants opted for verbal description as the best way of rendering sounds. As for emotions, most people preferred not to have any description, stating that emotions can be easily inferred from the image and context.

All in all, we need to note that in the case of this particular research study it was difficult to draw general conclusions because a large amount of eye-tracking data, comprehension scores and preference tests were frequently contradictory. For example, viewers tended to prefer other options than those which yielded highest comprehension scores or best eye-tracking results. What is more, viewers in different countries had different opinions as to their preferred options. Since there were many questions that remained unanswered, we decided to carry out another eye-tracking study on subtitling for the deaf and hard of hearing, this time focusing in greater detail on the question of reading speed, the editing of subtitles and differences in reading between various film genres.

4. Subtitling for the deaf and hard of hearing on digital television[20]

The goal of this on-going study is to examine the reading process of verbatim and edited subtitles among people with hearing impairments acquired at the pre-, peri- and post-lingually as well as among hearing people (the control group). The study material consists of both Polish and foreign productions. The genres examined include feature films, TV series, documentaries and news programmes.

4.1. Rationale

Since the previous study, conducted under the auspices of DTV4ALL, did not produce conclusive results on some of the most important issues related to SDH, we decided to investigate them further in more detail. The present study focuses on the questions of optimal reading speed (the two speeds tested are 12 vs. 15 cps[21]), subtitle editing (verbatim vs. edited), the influence of the film language and the film genre on the process of reading subtitles by hearing, hard of hearing and deaf viewers. Other research questions we intend to examine include the reading of content vs. grammar words, the influence of line segmentation on reading patterns, and the influence of shot changes on the re-reading of subtitles.

4.2. Participants

So far we have tested over 130 people with various degrees of hearing impairment as well as hearing people. Since this is an on-going study, the exact data on all the participants will be reported elsewhere.

4.3. Stimuli

Participants watch one of the two sets of clips (see Table 3) from five groups of materials: Polish feature films/ TV series, foreign[22] feature films, Polish documentary, foreign documentary and Polish news programmes. The two sets differ in the reading speeds of subtitles; each clip is displayed with the speed of either 12

20 This study is supported by research grant No. IP2011 053471 from the Polish Ministry of Science and Higher Education for the years 2011-2013.

21 Characters per second

22 'Foreign' here means English-speaking.

cps or 15 cps. The clips subtitled at the speed of 15 cps contain a verbatim version of the dialogue while those subtitled at the speed of 12 cps are edited, i.e. some linguistic structures have been simplified and elements of spoken language removed.

	Version 1	Version 2
1. **Polish feature film**		
Bitwa warszawska 1920 r.	12 cps	15 cps
Sami swoi	15 cps	12 cps
Londyńczycy	12 cps	15 cps
2. **Foreign feature film**		
Annie Hall	15 cps	12 cps
King's speech	12 cps	15 cps
Love actually	15 cps	12 cps
3. **Polish documentary**		
Największe cuda przyrody[23]	12 cps	15 cps
Polskie państwo podziemne	15 cps	12 cps
4. **Foreign documentary**		
Super Size Me	15 cps	12 cps
Roman Polański: Wanted and Desired	12 cps	15 cps
5. **News programme**		
Teleexpress	15 cps	12 cps
Wiadomości	12 cps	15 cps
Fakty	15 cps	12 cps

Table 3. Two sets of subtitled material

4.4. Procedure

Eye movements are recorded with SMI Red at the sampling frequency of 120 Hz. Participants are seated in front of a monitor at a distance of ca. 60 cm. After signing an informed consent to participate in the study, the participants answer a few personal questions displayed on the monitor, regarding their age, degree of

23 The original language of the film is English and its original title is *Great Natural Wonders of the World.* It is treated here, however, as a Polish film because the original voice was removed and only the Polish voice artist can be heard.

hearing loss, hearing aid/ implant, education, language of daily communication, and fluency in English.

As this is an on-going study, its results will be reported in future publications.

5. Eye-tracking and audio description

Apart from using eye-tracking to research subtitling for the deaf and hard of hearing, we have also examined its application to audio description. AD is a special description of what is taking place on the screen, originally created for blind and partially sighted viewers. Our recent studies have demonstrated a great potential of using AD in the education of blind and partially sighted children (A. Walczak 2010, A. Walczak, A. Szarkowska (forthcoming)). Taking advantage of eye-tracking, we also studied the applicability of AD to the education of sighted children (I. Krejtz et. al 2012, K. Krejtz et al. 2012). In the remaining part of the article we first examine previous eye-tracking studies on AD and then report on the results of our eye-tracking research on AD in the education of sighted children (for more details see I. Krejtz et. al 2012, K. Krejtz et al. 2012) and AD in arts.

While there has been quite a number of using eye-tracking methodology to study subtitling, there are not so many on eye-tracking and audio description. A large-scale research on AD was carried out within the framework of the European DTV4ALL project (see A. Chmiel, I. Mazur 2012, I. Mazur, A. Chmiel 2012). The goal of the project was to establish European AD guidelines and addressed cross-cultural and cross-linguistic differences in various EU states. Most of the study was based on the Pear Tree Project (see the Special Issue of *Perspectives: Studies in Translatology* vol. 20 no.1, 2012).

Another smaller eye-tracking study was conducted by the team involved in DTV4ALL. It examined perceptions of sighted students from Poland, Italy and Spain watching short video segments from *Marie Antoinette* and then compared them with the existing English AD (see A. Chmiel et al. 2010). It turned out that the participants' gaze did not always coincide with the AD script. I. Mazur and A. Chmiel continue looking into *Marie Antoinette* in their own project on 'Eye-tracking in audio description: perception of sighted viewers and its reflection in film descriptions for the blind', funded by the Polish Ministry of Science and Higher Education (see I. Mazur, A. Chmiel 2011).

A recent study by A. Vilaró et al. (2012) addressed the question whether – and if so, how – the soundtrack influences the perception and understanding of audiovisual texts. In their experiment, they tested four different soundtracks to the same clip (*Pear Film*), comparing differences in scanpaths between viewers.

The study demonstrated that the addition of sound does influence visual perception and eye movements.

Vilaró and Orero (forthcoming) examined the topic of leifmotifs, or recurrent themes, in audio description. They compared eye-tracking data with comprehension tests of a group of sighted volunteers who watched film fragments with and without AD. As predicted, people who watched the clips without AD did not notice an important detail (here a halo over the cockerel's head in the Pathé logo), as opposed to those who watched the audio described clip. A similar experiment was conducted by Orero and Vilaró (forthcoming), who looked into the question whether minute but important details regarding film characters match the participants' eye gaze and should be audio described. Participants watched three film excerpts, which were followed by questionnaires. The results showed an agreement between where the participants looked and what was audio described.

In a study on making meaning in AVT, J.-L. Kruger (2012: 67) attempted to 'correlate eye-tracking data […] with viewer constructions of the narrative'. Basing on recordings of eye movement data from Wallace Chafe's 1975 *Pear Film* combined with post-test written accounts, Kruger examined how viewers 'make sense of narrative sequences' (*ibid.*). In the eye-tracking experiment, he analysed dwell time, glance count and fixation count on objects marked with dynamic AOIs, i.e. adjusting the shape and size of the objects every few frames along with their movements in the film changing over time. The findings of the experiment suggest that 'visually peripheral elements that play a covert, top-down role in the narrative (i.e. with a higher degree of narrative salience) gain particular narrative importance when competing with the more overt, bottom-up aspects of the narrative (with an equally high narrative salience, as well as a high visual salience)' (J.-L. Kruger 2012: 67).

6. Audio description in education

Audio description is usually targeted at people who are blind or visually impaired. These viewers, as demonstrated by a number of previous studies, greatly benefit from the addition of AD to audiovisual materials. For example, an interesting study by G. Frazier and I. Coutinhoo-Johnson (1995) compared the comprehension of a video among visually impaired and sighted students. It was found that the visually impaired students who watched the audio described video had similar comprehension scores to sighted students and significantly better scores to those visually impaired students who saw the video without AD. This finding was corroborated by E. Peli, E.N. Fine and A.T. Labianca (1996), who found that visually impaired participants watching an audio described video performed better in comprehension tests than sighted participants and those visually impaired who saw the video without AD. Another experiment by E. Schmeidler and C. Kirchner (2001), which tested comprehension among blind people watching a scientific programme with and without AD, confirmed that those participants who watched the programme with AD retained more information than those who did not hear audio description.

6.1 Rationale

The research on the possibility of adding AD to educational materials was conducted by A. Walczak (2010) and A. Walczak, A. Szarkowska (forthcoming), demonstrating promising results. Our recent studies have shown that blind and visually impaired children can benefit from audio described material. Encouraged by these findings, we decided to examine the plausibility of using AD in the education of sighted children. We hypothesised that AD can foster children's comprehension and guide their attention to what is most relevant on screen. AD could thus be used as another educational tool and a supplement to traditional education.

When watching static images, people first tend to explore the image with long saccades and shorter fixations; after a few seconds the saccade amplitude drops and fixations become longer (see D.E. Irwin, G.J. Zelinsky 2002, B.M. Velichkovsky et al. 2005). This pattern can be attributed to two modes of acquiring information: exploration and inspection. The former is used to explore the visual field in order to select a target, and the latter to inspect the target in greater detail. This division corresponds to what has been known in literature as the distinction between focal-ambient (C.B. Trevarthen 1968). When watching dynamic scenes, such as films, the two modes may interplay with each other as scenes change over time.

In this study, we hypothesised that the addition of AD to the film will result in modifying children's eye movements, guiding children's visual attention, which would be reflected in a larger number of fixations, longer fixations and shorter saccades. This could indicate a more focused watching experience, resulting in better comprehension of educational content.

6.2. Participants

We tested a group of 44 primary school children aged 8-9, who watched an educational film with and without AD. Owing to poor calibration results, data from three participants were removed, giving *the* total of 41 participants (21 girls and 20 boys). Written consents to take part in the study were obtained from both the parents and the children. After the test, each child was rewarded with a Lego figure.

6.3. Stimuli

Children watched two 2-minute videos from the episode 'Blood' from the educational series *Once upon a time… Blood*. The experimental group watched the videos with text-to-speech audio description (for more on TTS AD see A. Szarkowska 2011).

There were two types of audio description used in the clips: synchronised and foreshadowed. Synchronised AD is the one that appears simultaneously with the action taking place on the screen. Foreshadowed AD precedes the action because when the action takes place, it is impossible to insert the description owing to the presence of dialogue (the major rule in AD is not to talk over the dialogue).

6.4. Procedure

Before the tests, the children made a lab tour and were shown the eye-tracker. They could also see how their eyes were monitored on the experimenter's screen. During the tests, children's eye movements were recorded with SR Research Ltd.'s Eye-Link CL eye-tracking system at 250 Hz. Children were seated at a distance of ca. 60 cm in front of a 17-inch LCD monitor.

Given that we were testing children and that we wanted the viewing to resemble natural conditions, we decided against using a chinrest. In consequence, we used the eye-tracker's remote mode of tracking, which requires adhesion of a contrasting dot to the participant's forehead to correct for their head movements. To facilitate calibration, we asked the children to 'shoot with their eyes' at the dots appearing on the screen.

Children were randomly assigned to two groups watching video with (the experimental group) or without AD (the control group). The experimental group was shown two 2-minute clips with AD, whereas the control group watched the same clips without AD. After watching each clip, children were presented with a scene recognition test: we showed them five pairs of pictures, where in each pair one picture was taken from the fragment they had seen and the other from another fragment of the film. Children's task was to indicate which picture they had seen in the film.

6.5. Results

In analysing data, we looked at fixation duration, fixation count and saccade amplitude. We also observed differences in the two types of AD (synchronised and foreshadowed).

A three-way mixed ANOVA analysis showed significant interaction effect between the experimental/ control group, synchronised/ foreshadowed AD and gender. In the experimental group during foreshadowed AD mean fixation duration was significantly longer (*M=413.04, SD=117.85*) than during synchronised AD (*M=356.04, SD=68.57*). Longer fixations may have resulted from listening to AD and anticipating the action about to happen. No significant differences in the control group were observed.

Another significant interaction was that between group and gender: it was boys' eye movements that were more influenced by AD than girls'. When watching the audio described clips, boys from the experimental group had longer fixations than boys from the control group. This may demonstrate a more focal attentional processing.

In order to verify whether children were actually looking at the parts of the video being described in AD, we also calculated fixation count in the specific areas of interest (AOI). For instance, one AOI was drawn around viruses which appeared in AD and were signalled by AD as follows: 'Suddenly, yellow crawling creatures – viruses – appear'. We analysed the period of two seconds starting with the moment the viruses appeared on screen (see Fig. 5).

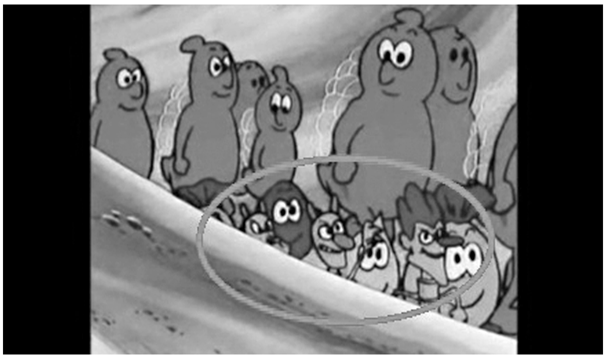


Fig. 5 Area of interest on viruses described in AD

In line with our predictions, it turned out that children in the experimental group had significantly more fixations on viruses (*M = 4.0, SD = 1.22*) than children from the control group (*M = 3.13, SD = 1.34*). This confirmed our hypothesis that AD guides children's attention to relevant elements of the scenes. As explained by Krejtz I. et. al. (2012), 'children followed virus movements more closely once a meaning was assigned to their creature representations'.

After eye-tracking tests, we asked the children to do a scene recognition test, where they had to choose which picture from each pair appeared in the clip they had seen (see Fig. 6).

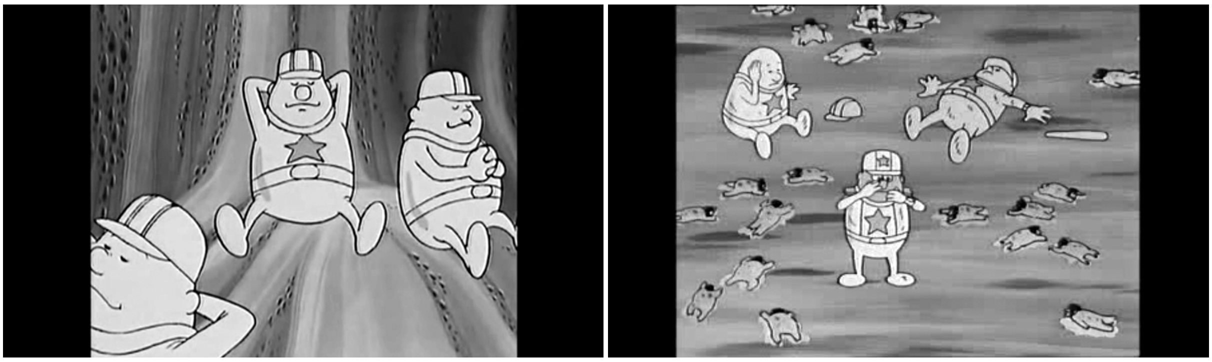


Fig. 6 A sample pair of pictures from the scene recognition test

Both groups reached nearly 100% correctness on this task and no significant differences were observed between conditions. However, when analysing answers to open-ended questions asked after the test during individual debriefing, we observed interesting results regarding the use of specialised vocabulary. The use of specialised terms was tested by asking children about mitosis, the role of antibodies and who attacked the organism (see Fig. 7).

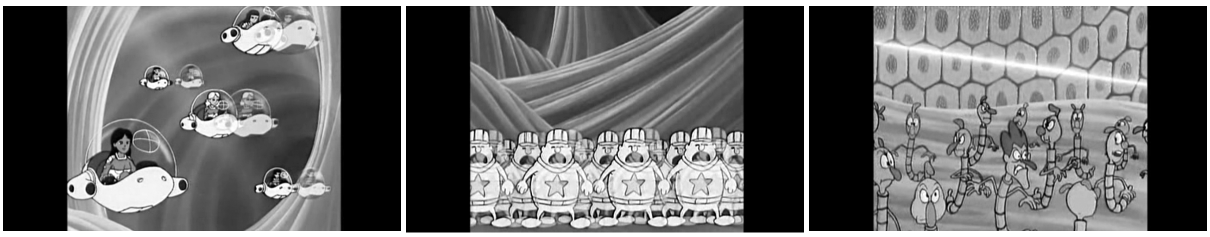


Fig. 7: Film stills showing the mitosis of white cells (picture 1 and 2) and viruses who attached the organism (picture 3)

Their answers were coded on a 5-point scale, where 0 stood for none, 1: wrong, 2: poor, 3: good, 4: very good. Very few children said they couldn't remember the answers or who gave incorrect answers (23.8% in the control group and 14.3% in the AD group). We then examined the answers ranked as good or very good, the difference being whether a specialised term was used or not. There was a significant correlation between the group (experimental/ control) and the answer. Over 70% of children from the experimental group used specialised terms as opposed to nearly 30% of children from the control group. Table 4 presents samples of debriefing interviews, one from the experimental group and one from the control group[24].

Experimental group	Control group
R: OK, tell me now, do you remember who copied themselves… got replicated? *G: White cells and these… ships.* *R: Alright. And what did antibodies do?* *G: Antibodies did… they… scared away the… they crawled over the viruses and as if they started to melt, something like that.* *R: OK, now, and who attacked the organism?* *G: Viruses.* *R: Great, and what is mitosis?* *G: Dividing into twos.* *R: Fantastic, thank you very much.*	*R: Alright, tell me if you remember who copied themselves?* *G: White cells and girls in ships copied themselves.* *R: Great, and what do antibodies do?* *G: Antibodies... which ones were antibodies?* *R: Well, what do you remember… do you remember which of them were the antibodies?* *G: No.* *R: OK, it doesn't matter. And do you remember who attacked the organism?* *G: The organism was attacked by the… the sticks.* *R: Uhm.* *G: The brown sticks, I don't know what they're called.* *R: Ok, it's alright. And what is mitosis?* *G: I have no clue.* *R: Good, that's all.*

Table 4. Sample post-eye-tracking test interviews with two girls: (1) from the experimental and (2) the control group, R =researcher, G =girl

To sum up, the findings of our study show that children's eye movements – and thus their attention – can be positively affected by AD, which helps them to focus on relevant elements of the screen. The AD script also provides children with specialised vocabulary presented in context, which fosters their understanding of educational material and facilities knowledge acquisition.

7. Audio description in art

Apart from standard audio guides for sighted viewers, more and more museums and art galleries are offering audio description for the blind and partially sighted visitors. Such description usually includes standard information like the name of the artist, the title of the painting/ artwork, date, dimensions (height/ width), technique as well as a verbal description of the visual aspects of the work. Following the tenets of universal design theory, according to which a product/

24 Translated from Polish by A.Sz.

service should be designed from conception to include as many potential and differently abled users as possible (see J.-P. Udo, D.I. Fels 2009: 1), 'exhibit designers are increasingly combining standard audio tours with audio descriptions, an 'all-in-one' concept' (ADP Standards 2009: 28). In consequence, both sighted and visually impaired people can receive the same audio description. As reported in ADP Standards (*ibid.*), sighted visitors also benefit from extended verbal descriptions of artworks. Such audio description can also be used in education as an aid to visual literacy.

This new application of audio description poses new research questions and opens new research avenues. Below we look at one such area regarding the perception of artworks – with and without audio description – by sighted students. The aim of the study was to examine the influence of audio description on perception of art among sighted young people.

Our hypothesis was that AD can precisely guide our attention to the most relevant parts of the painting enabling better understanding of an artwork and fostering visual memory of the painting. AD can thus help viewers acquire new information such as specialized history about particular piece of art, promoting more elaborated information processing and stimulating their visual memory.

7.1. Participants

All participants (N=59) were students of the first two years of high school profiled towards mathematics and computer science. There were 22 male and 27 female participants.

7.2. Experimental design

The present research was a 3 (type of presentation) x 2 (order of presentation) experiment. The type of presentation was a between-subject factor. Each participant was randomly assigned to one of the three experimental conditions (types of presentation): (1) the AD group: visualization of art piece with auditory description accompanying the picture, (2) the text group: visualization of art piece with textual description presented next to the picture, (3) the control group: visualization of art piece without any description. The order of presentation was a within-subject factor. Each participant saw two paintings in random order.

7.3. Stimuli

Participants were shown two short video clips each with a still picture of two famous paintings: *The Umbrellas* by Pierre-August Renoir and *The calling of St. Matthew* by Caravaggio. The first clip lasted for 4 min and 18 sec. and the second one for 4 min. 35 sec.

In all experimental conditions we presented the students with a still image of the painting situated on the left-hand side of the screen. The experimental conditions differed in what was presented on the right-hand side of the screen. In the AD condition, the AD script was read by a voice talent (previously recorded in a studio), so the participants did not see the text, and the space on the right of the screen was left blank (see Fig. 8a). In the text condition, the script was presented in the textual form next to the picture (see Fig. 8b). In the control condition, the space on the right was again left empty (see Fig. 8c).

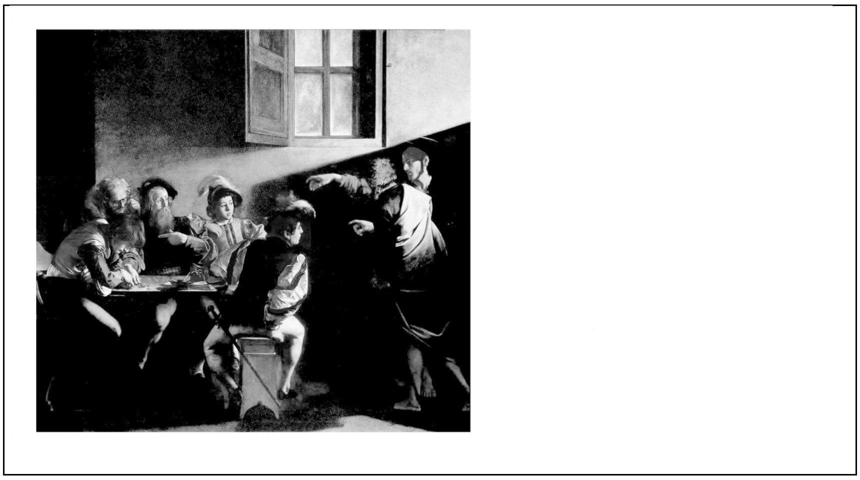


Fig. 8a. A sample scene from the video clip presenting the Caravaggio painting in the AD condition

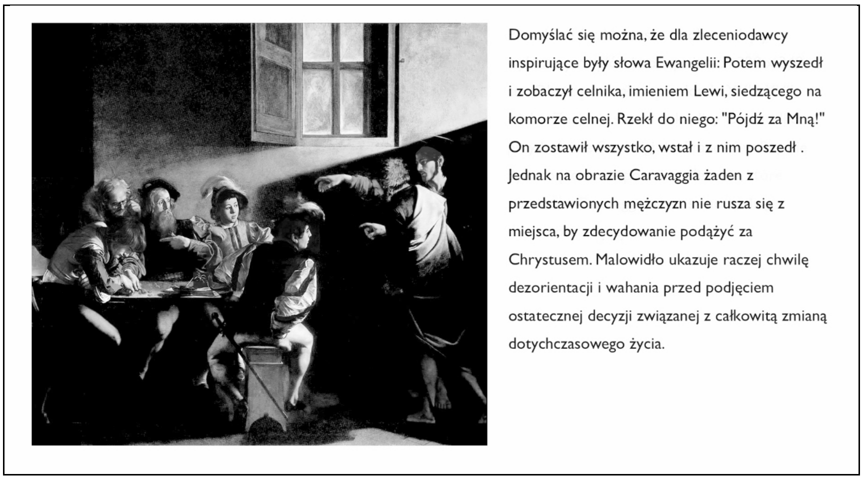


Fig. 8b: A sample scene from the video clip presenting the Caravaggio painting in the text condition

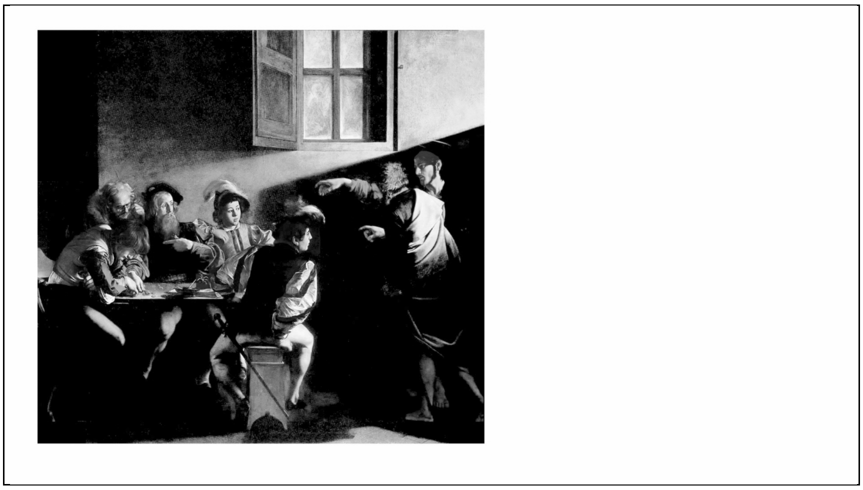


Fig. 8c: A sample scene from the video clip presenting the Caravaggio painting in the control condition

7.4. Apparatus

Eye movements were recorded at 250 Hz with the SMI eye-tracking system. Participants were seated in front of a computer monitor (1024 x 768 resolution; 22-inch LCD, 60 Hz refresh rate). The eye-tracker's remote mode of tracking was used to promote participants' comfort and ensure natural viewing conditions. The spatial resolution of used eye-tracking system was 0.03 degrees and gaze position accuracy 0.4 degrees. The SMI Experiment Center software was used to present stimuli and synchronize them with eye movements recorded. For fixations and saccades detection as well raw data cleaning, SMI BeGaze software was employed. We used default BeGaze settings for detecting fixations and saccades with the High Speed Event Detection method. With this method saccades are treated as primary events and fixations are derived from them. In order to capture saccades, the method uses a velocity-based algorithm. With this algorithm peak velocity threshold is set to 40 degrees per second and minimum saccade duration is set to 22 ms while minimum fixation duration is set to 50 ms.

7.5. Procedure

The experiment was conducted individually. First, each participant signed an informed consent; being underage, students were also obliged to provide a signed consent form from their parents. Then the 9-point calibration was performed. After a successful calibration procedure, two video clips with the paintings of Caravaggio and Renoir were presented in random order. After each video clip, participants were asked to solve interactive jigsaw puzzles[25] made from the painting they had just seen (see Fig. 9).

Fig. 9: Interactive jigsaw puzzle from Caravaggio's painting

After watching both video clips, participants answered a series of interactive questions on what they have learned from video clips about the painters and the works of art.

7.6. Results

A qualitative analysis of scanpaths in the AD and the control group clearly presents different viewing patterns (see Figures 10 and 11). As the AD group was closely following the audio description, they were more focused on the parts of the painting mentioned (e.g. the young man counting money at the table). In contrast, the control group is concentrated on faces, which is a typical viewing pattern, well-established in eye-tracking research, confirming that human face is a special stimulus always attracting our attention (see A.L. Yarbus 1967, O. Hershler, S. Hochstein 2005, S.R.H. Langton et al. 2008).

25 www.jigsawplanet.com

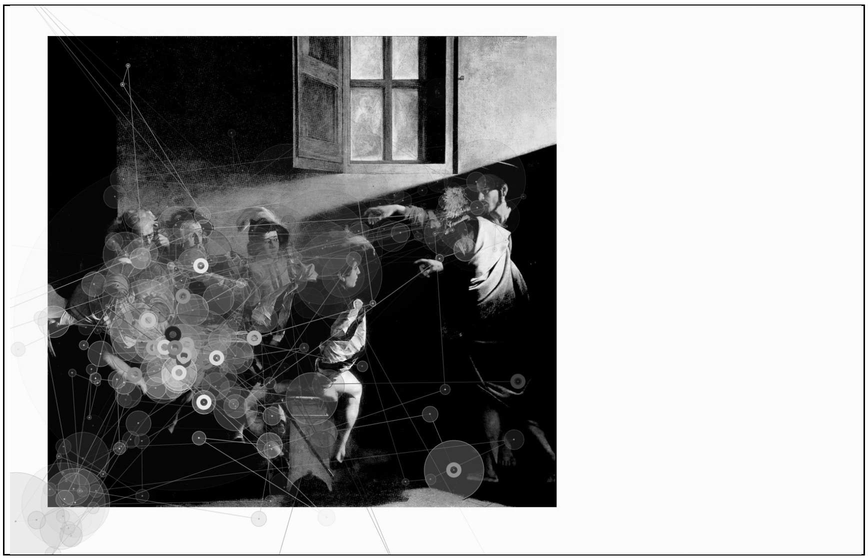


Fig. 10: *Two-second long scanpath in the AD group*

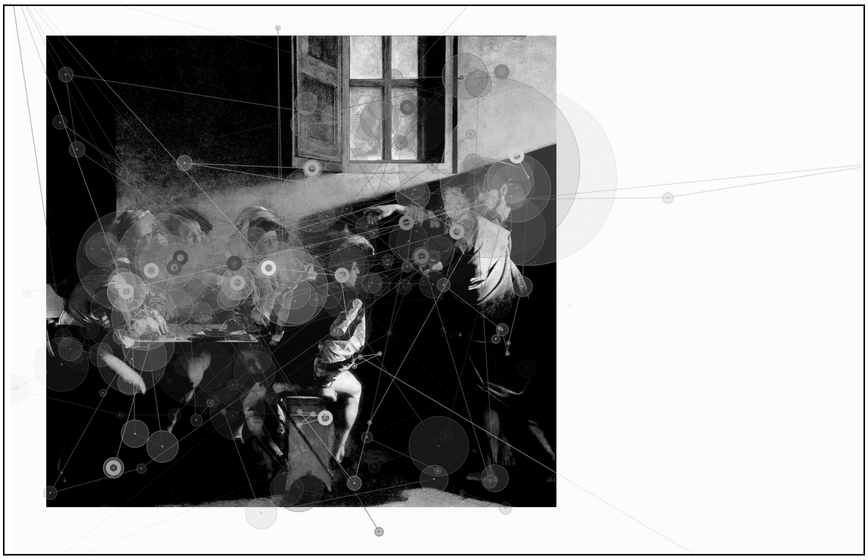


Fig. 11: *Two-second long scanpath in the control group*

To verify the influence of AD on the perception of paintings, we conducted a one-way between-subject analysis of variance, with group as the independent factor and percent of dwell time on the painting as a dependent variable. The analysis revealed a significant main effect of the experimental condition, $F(2,55) = 104.51$, $p < 0.001$. As expected, the AD group examined the painting significantly longer ($M = 76.91$; $SD = 10.73$) than the control group ($M = 60.99$, $SD = 9.73$) and the group with written text presented next of the painting ($M = 29.17$, $SD = 10.10$). The differences between all three conditions were statistically significant (see Fig. 16).

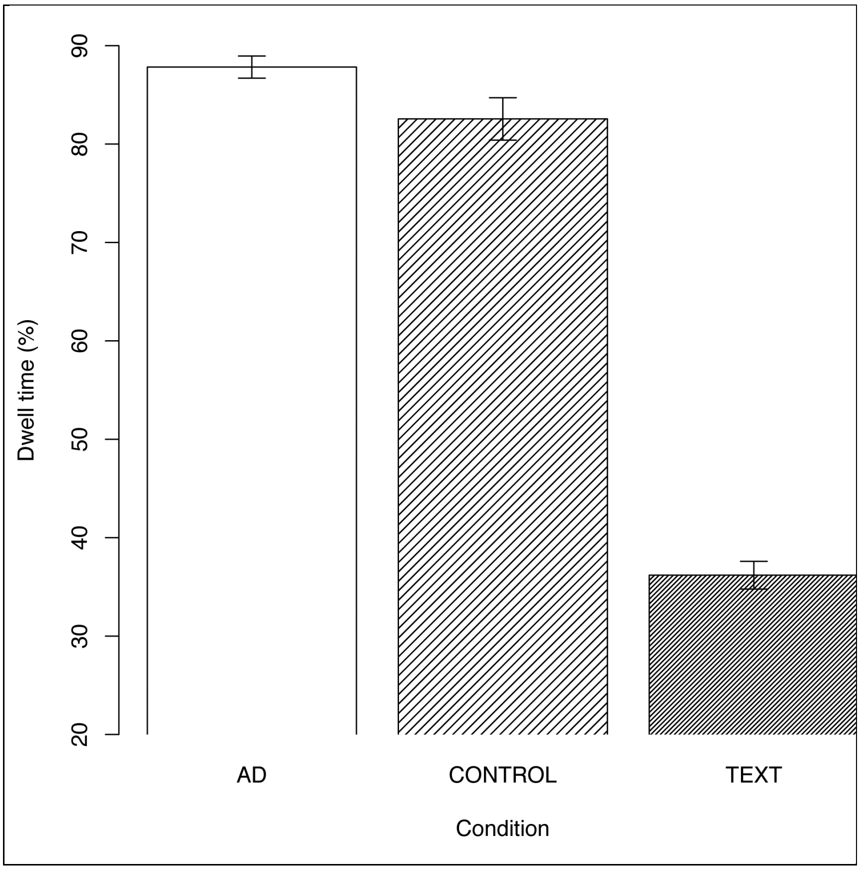


Fig. 12: Group differences in dwell time on the Caravaggio painting

These preliminary results are in line with our initial predictions: the AD group focused significantly longer on the painting. What is more, their eye movements followed the description of the painting read by the voice talent. This viewing pattern was markedly different from the other two conditions, which demonstrates that AD guides the viewer's attention towards the elements described in the painting.

We believe that AD can be an attractive aid to developing visual literacy. It can be used in museums, art galleries and educational settings as supplement information broadening young people's art perspectives. Because listening to AD while watching a piece of art employs two different modalities, it leaves some cognitive resources necessary for more thorough visual analysis.

8. Conclusions and future work

In this article we have presented four eye-tracking studies on subtitling for the deaf and hard of hearing and on audio description for the blind and partially sighted as well as sighted people. We hope to have shed some new light on how people watch subtitles and how they perceive audio described works. The results of our research can be implemented in the audiovisual translation professional market practices as well as in education.

Combined with other methods of data collection, such as preference questionnaires or comprehension tests, eye-tracking methodology has proved to be helpful in providing quantitative evidence to testify our hypotheses. However, the application of eye-tracking to audiovisual translation in general and to media accessibility in particular still remains a largely uncharted territory, haunted by the lack of established methodological tools. As rightly noted by J.-L. Kruger (2012: 71), so far most eye-tracking studies on AVT have analysed 'static elements in a dynamic scene and not […] dynamically changing areas of interest'. The application of eye-tracking tools designed for the analysis of static scenes seems inadequate because film is a dynamic medium. In a film, pictures change quickly over time, objects which are being fixated move and so do viewers' eyes. This requires, for instance, an analysis of smooth pursuits, which are not taken into consideration by most eye-tracking studies. The reason for this has naturally stemmed from the technical constraints offered by commercial software available until recently. However, with new software now available on the market it is possible to trace smooth pursuits as part of the dwell time (defined as the total duration of all fixations and saccades in a defined AOI, see J.-L. Kruger 2012) with high-speed eye-trackers.

References

American Council for the Blind (2009): *Audio description standards*. [www.acb.org/dp/docs/DP_Standards.doc].

Caffrey, C. (2012): *Eye-tracking application for measuring the effects of experimental subtitling procedures on viewer perception of subtitled AV content.* In: E. Perego (ed.), Eye-tracking and audiovisual translation. Roma. 225–260 (in press).

Carpenter, P.A./ M.A. Just (1983): *What your eyes do while your mind is reading*. In: K. Rayner (ed.), Eye movements in reading: perceptual and language processes. New York. 275–307.

Chmiel, A./ I. Mazur (2012): *Audio Description Made to Measure: Reflections on Interpretation in AD based on the Pear Tree Project Data*. In: A. Remael/ O. Pilar/ M. Carroll (eds), Audiovisual Translation and Media Accessibility at the Crossroads. Media for All 3. Amsterdam/ New York. 173–188.

Chmiel, A./ I. Mazur/ A. Vilaro/ E. Digiovanni/ P. Chmiel/ P. Orero (2010): *How Audio Description Influences Perception. A Work-In-Progress Report on an International Eye-Tracking Study*. ARSAD 2010.

D'Ydewalle, G./ W. De Bruycker (2007): *Eye Movements of Children and Adults While Reading Television Subtitles*. In: European Psychologist. 12(3). 196–205.

D'Ydewalle, G./ I. Gielen (1992): *Attention allocation with overlapping sound, image, and text*. In: K. Rayner (ed.), Eye movements and visual cognition: Scene perception and reading. New York. 415–427.

D'Ydewalle, G./ J. V. Rensbergen/ J. Pollet (1987): *Reading a message when the same message is available auditorily in another language: The case of subtitling*. In: J. K. O'Reagan/ A. Lévy-Schoen (eds), Eye Movements: From Physiology to Cognition. Amsterdam/ New York. 313–321.

de Linde, Z./ N. Kay (1999): *The semiotics of subtitling*. Manchester.

Duchowski, A. T. (2002): *A breadth-first survey of eye-tracking applications. Behavior Research Methods*. In: Instruments and Computers. 1–16.

d'Ydewalle, G./ C. Praet/ K. Verfaillie/ J. van Rensbergen (1991): *Watching Subtitled Television: Automatic Reading Behaviour*. In: Communication Research. 18(5). 650–666.

Frazier, G./ I. Coutinho-Johnson (1995): *The Effectiveness of Audio Description in Processing Access to Educational AV Media for Blind and Visually Impaired Students in High School*. San Francisco.

Ghia, E. (2012): *The impact of translation strategies on subtitle reading*. In: E. Perego (ed.), Eye-tracking and audiovisual translation. Roma. 155-181.

Hershler, O./ S. Hochstein (2005): *At first sight: A high-level pop-out effect for faces*. In: Vision Research. 45. 1707–1724 (in press).

Irwin, D. E./ G. J. Zelinsky (2002): *Eye Movements And Scene Perception: Memory For Things observed*. In: Perception and Psychophysics. 64. 882–895.

Jensema, C. (1998): *Viewer reaction to different television captioning speeds*. In: American Annals of the Deaf. 143. 318–324.

Jensema, C./ R.S. Danturthi/ R. Burch (2000b): *Time spent viewing captions on television programs*. In: American Annals of the Deaf. 145(5). 464–468.

Jensema, C./ S.E. Sharkawy/ R.S. Danturthi/ R. Burch/ D. Hsu (2000a): *Eye movement patterns of captioned television viewers*. In: American Annals of the Deaf. 145(3). 275–285.

Krejtz, I./ A. Szarkowska/ A. Walczak/ K. Krejtz/ A. Duchowski (2012): *Audio Description as an Aural Guide of Children's Visual Attention: Evidence from an Eye-Tracking Study*. In: ETRA 2012 Proceedings of the Symposium on Eye-tracking Research and Applications, pp. 99-106. ACM New York, NY. [http://dl.acm.org/citation.cfm?doid=2168556.2168572].

Krejtz, K./ I. Krejtz/ A. Duchowski/ A. Szarkowska/ A. Walczak (2012): *Multimodal learning with audio description: an eye-tracking study of children's gaze during a visual recognition task*. In: Proceedings of the ACM Symposium on Applied Perception (SAP '12): ACM, New York, NY, USA, 83-90. DOI=10.1145/ 2338676.2338694

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

Langton, S. R. H./ A. S. Law/ A. M. Burton/ S. R. Schweinberger (2008): *Attention capture by faces*. In: Cognition. 107. 330–342.

Majaranta, P./ K.-J. Raiha (2002): *Twenty Years of Eye Typing:Systems and Design Issues*. In: Eye-tracking Research & Applications (ETRA) Symposium. New Orleans, LA.

Mazur, I./ A. Chmiel (2011): *Odzwierciedlenie percepcji osób widzących w opisie dla osób niewidomych. Badania okulograficzne nad audiodeskrypcją*. In: Lingwistyka Stosowana/ Applied Linguistics/ Angewandte Linguistik. 4. 163–180.

Mazur, I./ A. Chmiel (2012): *Towards common European audio description guidelines: results of the Pear Tree Project* In: Perspectives: Studies in Translatology. 20/1. 5-23.

Moran, S. (2012): *The effect of linguistic variation on subtitle reception*. In: E. Perego (ed.), Eye-tracking and audiovisual translation. Roma. 183–223 (in press),

Orero, P./ A. Chmiel (forthcoming): *Eye-tracking analysis of minor details in films for audio description*. In: R. Agost/ E. di Giovanni/ P. Orero (eds), Multidisciplinarity in audiovisual translation [Special issue]. MonTi. 4.

Peli, E./ E. M. Fine/ A. T. Labianca (1996): *Evaluating Information Provided By Audio Description*. In: Journal of Visual Impairment and Blindness. 90. 378–385.

Perego, E./ F. de Missier/ M. Porta/ M. Mosconi (2010): *The cognitive effectiveness of subtitle processing*. In: Media Psychology, 13, 243–272.

Rajendran, D. J./ A.T. Duchowski/ P. Orero/ J. Martínez/ P. Romero Fresco (2011): *Effects of text chunking on subtitling: A quantitative and qualitative examination*. [www.andrewd.ces.clemson.edu/ research/ vislab/ docs/ perspectives.pdf].

Rayner, K./ S.A. Duffy (1988): *On-line comprehension processes and eye movements in reading*. In: M. Daneman/ G.E. MacKinnon/ T.G. Waller (eds), Reading research: Advances in theory and practice. New York. 13–66.

Rayner, K. (1998): *Eye movements in reading and information processing: Twenty years of research*. In: Psychological Bulletin. 124. 372–422.

Rayner, K./ A. Duffy/ J. Ashby/ Ch. Clifton, Jr (2012): *Psychology of reading, 2nd edition*. New York.

Schmeidler, E./ C. Kirchner (2001): *Adding Audio Description: Does It Make a Difference?* In: Journal of Visual Impairment and Blindness. 95(4). 198–212.

Szarkowska, A. (2010): *Accessibility to the Media by Hearing Impaired Audiences in Poland: Problems, Paradoxes, Perspectives*. In: J. Díaz Cintas/ A. Matamala/ J. Neves (eds), New Insights into Audiovisual Translation and Media Accessibility. Media for All 2. Rodopi: Amsterdam/ New York. 139–158.

Szarkowska, A. (2011): *Text-to-speech audio description: towards wider availability of AD*. In: Journal of Specialised Translation. 15. 142–163.

Szarkowska, A./ I. Krejtz/ Z. Kłyszejko/ A. Wieczorek (2011): *Verbatim, standard, or edited? Reading patterns of different captioning styles among deaf, hard of hearing, and hearing viewers*. In: American Annals of the Deaf. 156 (4). 363–378.

Trevarthen, C.B. (1968): *Two mechanisms of vision in primates*. In: Psychologische Forschung. 31. 229–337.

Udo, J.-P./ D.I. Fels (2009): *The Rogue Poster-Children of Universal Design: Closed Captioning and Audio Description*. In: Ted Rogers School of Information Technology Management Publications and Research. Paper 18. http:/ / digitalcommons.ryerson.ca/ trsitm/18.

Velichkovsky, B.M./ M. Fels/ J.R. Helmert/ S. Pannasch (2005): *Two Visual Systems and Their Eye Movements: Evidence from Static and Dynamic Scene Perception*. In: CogSci 2005: Proceedings of the XXVII Conference of the Cognitive Science Society. 2283–2288.

Vilaró, A./ P. Orero (forthcoming): *The audio description of leitmotifs*. In: S. Bruti/ E. di Giovanni/ P. Orero (eds), Audio visual translation across Europe: An ever-changing landscape. Bern/ Berlin.

Vilaró, A./ A. Duchowski/ P. Orero/ T. Grindinger/ S. Tetreault/ E. di Giovanni (2012): *How Sound is The Pear Tree Story? Testing the Effect of Varying Audio Stimuli on Visual Attention Distribution*. In: Perspectives: Studies in Translatology 20 (1): Special Issue on The Pear Stories. 55–65.

Walczak, A. (2010): *Audio description for children. A case study of text-to-speech audio description of educational animation series Once Upon a Time... Life* (unpublished MA thesis) [http:/ / avt.ils.uw.edu.pl/ files/ 2012/ 09/ Walczak_TTSAD-for-children.pdf].

Walczak, A./ A. Szarkowska (forthcoming): *Text-to-speech audio description. Testing educational materials with visually-impaired children*.

Yarbus, A. L. (1967): *Eye movements and vision*. New York.