The Cognitive Effectiveness of Subtitle Processing

ELISA PEREGO
Department of Language, Translation and Interpreting Studies, University of Trieste, Trieste, Italy

FABIO DEL MISSIER
Department of Psychology, University of Trieste, Trieste, Italy

MARCO PORTA and MAURO MOSCONI
Department of Computer and Systems Engineering, University of Pavia, Italy

In an experimental study, we analyzed the cognitive processing of a subtitled film excerpt by adopting a methodological approach based on the integration of a variety of measures: eye-movement data, word recognition, and visual scene recognition. We tested the hypothesis that the processing of subtitled films is cognitively effective: It leads to a good understanding of film content without requiring a significant tradeoff between image processing and text processing. Following indications in the psycholinguistic literature, we also tested the hypothesis that two-line subtitles whose segmentation is syntactically incoherent can have a disruptive effect on information processing and recognition performance. The results highlighted the effectiveness of subtitle processing: Regardless of the quality of line segmentation, participants had a good understanding of the film content, they achieved good levels of performance in both word and scene recognition, and no tradeoff between text and image processing was detected. Eye-movement analyses enabled a further characterization of cognitive processing during subtitled film viewing. This article discusses the theoretical implications of the findings for both subtitling and multiple-source communication and highlights their methodological and applied implications.

Address correspondence to Elisa Perego, Department of Language, Translation and Interpreting Studies, University of Trieste, Via Filzi, 14, Trieste 34132, Italy. E-mail: eperego@units.it
INTRODUCTION

The main aim of the present article is to test theoretically grounded hypotheses on the processing of subtitled films by analyzing the relationship between attention allocation, word recognition, and scene recognition. In particular, we tested experimentally the hypothesis that processing a subtitled product is cognitively effective, that is, it leads to a good understanding of the film content while not requiring a tradeoff between image processing and text processing. Another specific aim of our study was to test experimentally the influence of line segmentation quality in two-line subtitles on cognitive processing. Unlike previous research, our experiment adopted an integrated approach, taking into account eye movements (a proxy for attention allocation), gist comprehension, recognition of specific words contained in the subtitles (outcome measure of subtitle processing), and recognition of film scenes (outcome measure of visual scene processing). This approach enabled us to obtain a richer picture of the cognitive processing of subtitled films. The results of the study raised various theoretical, methodological, and applied implications.

In this introductory section, a brief review of relevant findings on subtitle processing will be provided. Afterward, the main hypotheses of our experiment on cognitive effectiveness and subtitle line segmentation will be put forward, followed by a description of the experiment, and by the presentation of recognition results and eye-movement analyses. Finally, the theoretical, methodological, and applied implications of the findings will be discussed.

The Cognitive Processing of Subtitles: An Underinvestigated Topic

Subtitling is a multiple-purpose and versatile solution that has been used for a long time now in various audiovisual contexts with different aims (Díaz Cintas, 2001; Díaz Cintas & Remael, 2007; Luyken, Herbst, Langham-Brown, Reid, & Spinhof, 1991). First, subtitling is commonly used as a low-cost translation method to make audiovisual products available to various linguistic communities without resorting to dubbing (e.g., Koolstra, Peeters, & Spinhof, 2002). Subtitling is indeed much less expensive than dubbing and it may be preferred by individuals whose competence in the film's original language enables them to enjoy the original voices of the actors and the original “atmosphere” of the product. Subtitling has also been used in instructional contexts as a tool to promote second language learning, with a proven advantage for lexical acquisition (e.g., d’Ydewalle & Pavakanun, 1995, 1997; d’Ydewalle & Van de Poel, 1999). Finally, subtitling in the same language of the original soundtrack is commonly employed to support the deaf and hard of hearing (de Linde & Kay, 1999a, 1999b) or, in some cases, to
promote literacy (Kothari, 1998, 2000; Kothari & Takeda, 2000). In spite of its usefulness, social relevance, and cost effectiveness, subtitling has received little attention from experimental psychology and communication studies. For this reason, practical applications of subtitling rely more on conventional rules of thumb than on validated empirical guidelines. Research in this area can, therefore, have significant applied value.

Research on subtitling can also be interesting from a theoretical point of view, being a natural case of multiple-source processing and a good example of information integration in communication (e.g., d’Ydewalle & De Bruycker, 2007; Fox, Park & Lang, 2007; Gielen, 1988; Lang, 1995, 2000; Paivio, 1971, 1986; Zhou, 2004). In the case of subtitled films, the individual receives visual input from the images and the text of the subtitles, and verbal and nonverbal auditory information from the soundtrack. Understanding how individuals are able to process, integrate, and understand information coming from these different sources will indeed provide a significant contribution to the theoretical understanding of complex communication processes. Research on subtitles can thus be helpful to understand other forms of multiple-source communication in which the individual has to pay attention to information delivered through various sensorial channels and information sources (e.g., watching and hearing the news on television while processing the headlines and/or other written information on the screen).

A Short Review of Previous Findings

Although earlier studies have provided valuable information on the cognitive processing of subtitled material, the picture is still partial (for reviews, see de Linde & Kay, 1999a, 1999b; d’Ydewalle & Gielen, 1992; d’Ydewalle & Pavakanun, 1995, 1997; d’Ydewalle, Muylle, & Van Rensbergen, 1985; d’Ydewalle & Van Rensbergen, 1989; Koolstra & Beentjes, 1999; Koolstra, Van der Voort, & d’Ydewalle, 1999; Praet, Verfaille, De Graef, Van Rensbergen, & d’Ydewalle, 1990). A first finding is that individuals show a rather strong tendency to initiate reading subtitles at their onset even when they have little experience with this source of information (d’Ydewalle & Gielen, 1992; d’Ydewalle & Pavakanun, 1997; d’Ydewalle, Praet, Verfaille, & Van Rensbergen, 1991; d’Ydewalle, Van Resenbergen, & Pollet, 1987; Gielen, 1988), and this tendency is stronger when subtitles are informative (i.e., the language of the soundtrack is unknown or known little, there is limited redundancy between written text and images, and subtitles convey valuable information).

Moreover, people who have a degree of experience with subtitles seem to read them in a rather effortless way (e.g., d’Ydewalle & Gielen, 1992; see also d’Ydewalle & De Bruycker, 2007). The observation that reading text on screen seems to require relatively low effort is consistent with the generally agreed idea that reading can become partially automatic and rela-
tively effortless for most normal adult readers due to learning processes (e.g., LaBerge & Samuels, 1974; Logan, 1997; Perfetti, 1985; Perfetti & Marron, 1998). Indeed, although reading requires several complex perceptual and cognitive skills (e.g., knowledge of a given writing system, phonological awareness, decoding skills, grammatical expertise), once learned, readers appear to perform this activity without great effort, thanks to the partial automatization of various components of the task (e.g., LaBerge & Samuels, 1974; Logan, 1997; Rayner & Sereno, 1994; Stanovich, 1990).

d’Ydewalle and De Bruycker (2007) recently conducted a detailed study on the eye-movement patterns of children and adults. Eye movements were monitored while participants watched a foreign language program with subtitles. Although no clear differences in the eye-movement patterns were observed between children and adults, a more regular reading pattern was detected in standard two-line subtitles (vs. one-line subtitles). The authors concluded that two-line subtitles are more likely to be fully processed than one-line subtitles because they contain information that cannot be easily inferred from the visuals. The general conclusion was that participants behaved adaptively, paying more attention to nonredundant subtitles that are more useful to understand the program. Other interesting results from the same study were the observation of shifts between the visual scene and the subtitled area (which replicated previous findings), an increased frequency of regressive eye movements (vs. regular reading), and the shorter duration of fixations on subtitles (vs. regular reading). These results confirm the idea that viewers shift visual attention between different sources of information (subtitles and images) while watching a subtitled program. They also suggest that individuals may adjust fixation times and flexibly reconsider text that has already been processed in order to comply with particular processing needs and constraints of the situation.

The combination of the relatively effortless reading behaviour observed by scholars of the Belgian school (d’Ydewalle et al., 1987, 1991; d’Ydewalle & Van Rensbergen, 1989; Koolstra & Beentjes, 1999) and efficiency considerations (d’Ydewalle & De Bruycker, 2007; d’Ydewalle & Gielen, 1992) has been proposed as an explanation for most of the existing findings on subtitle reading. Unfortunately, the great majority of existing studies on the cognitive processing of subtitled material focused only on a limited subset of measures, emphasizing either attention allocation (eye-movement studies) or subtitle-related performance measures (e.g., subtitle recognition). The capacity to process and remember visual scenes, which has been widely studied in mass communication and redundancy research (e.g., Drew & Grimes, 1987; Grimes, 1991; Zhou, 2004), has been almost neglected in the field of audiovisual translation. In other words, studies on subtitle processing usually do not provide a comprehensive analysis of processing strategies and of their effectiveness, because they do not take into account at the same time eye movements, visual scene processing performance, and subtitle processing.
Cognitive Effectiveness of Subtitle Processing

As a consequence, studies simultaneously encompassing a larger set of measures are needed to properly appraise the effectiveness of subtitled material processing and to examine the potential tradeoff between scene processing and subtitle processing.

The Cognitive Effectiveness of Subtitle Processing

As pointed out by d’Ydewalle and De Bruycker (2007), a subtitled (translated) television program requires the processing of information coming from three different but potentially overlapping parallel sources: the visual image, the subtitles in the native language on the screen, and the soundtrack in a foreign language. From the perspective of early-selection theories of attention (Broadbent, 1958; Treisman, 1968), this task might be considered as rather taxing because it requires multiple attention shifts between text and images, the monitoring of at least two sources of complex information (subtitles and visual scenes) requiring different processing strategies, and the integration of information coming from these sources in order to reach a good understanding of the program (see also Bergen, Grimes, & Potter, 2005; d’Ydewalle & Van Rensbergen, 1989; Grimes, 1991; Lang, 1995, 2000; Paivio, 1971, 1986). Additionally, listening to the soundtrack in a completely unknown language may represent a potential source of interference coming from the audio channel (e.g., Banbury & Berry, 1998). If processing and integration of information from parallel sources is cognitively demanding, as suggested by early-selection theories of attention, then a tradeoff between visual processing and subtitle processing should be expected: When more cognitive resources are devoted to subtitle processing, visual scene processing will be less effective (and vice versa). Thus, a negative correlation between subtitle recognition measures and visual recognition measures should be observed.

However, given that reading and visual processing are relatively effortless and partially automatic activities per se (LaBerge & Samuels, 1974; Logan, 1997; Perfetti, 1985; Perfetti & Marron, 1998; Lang, 2001; Lang, Potter, & Bolls, 1999; Rayner & Pollatsek, 1989; Zhou, 2004) and that images often facilitate subtitle understanding by providing an appropriate context, the continuous integration of two visual sources (subtitles and visual scenes) might not be so complex as could be prima facie argued. Indeed, the un-interfering and sometimes even supporting role of images in the processing of media communicated messages has been highlighted by several theorists (Grimes, 1991; Lang, 2001; Zhou, 2004). The idea of cognitive effectiveness of subtitled material processing has been explicitly stated by d’Ydewalle and Gielen (1992): “When people watch television, the distribution of attention between different channels of information turns out to be an effortless process. Viewers seem to have developed a strategy that allows them to process these channels without problems and in which reading the subtitles occupies a major place” (p. 425). If this is the case, no tradeoff between
visual processing and subtitle processing should be observed in normal conditions, thanks to the efficiency of visual processing and reading, and to the development and application of smart strategies for attention allocation and information integration (but this might not hold in peculiar situations in which the information delivered through the various sources is complex and nonredundant, e.g., Drew & Grimes, 1987; Grimes, 1991).

Unfortunately, due to the abovementioned methodological limitations, previous research did not make convincingly clear whether processing a subtitled program can be handled with relatively low effort, achieving good levels of performance, and without causing a tradeoff between image processing and text processing (i.e., our cognitive effectiveness hypothesis). Given this state of the art, the first goal of our study was to appraise the cognitive effectiveness of subtitled film processing. To this aim we assessed the existence of a tradeoff between scene and subtitle processing, analyzing the participants’ ability to recognize specific words presented in the subtitles and freeze-frames belonging to the film fragment they had been exposed to. Furthermore, to gain a better understanding of subtitled material processing, we resorted to the analysis of eye movements which enabled us to trace the direction of visual attention and to make some inferences on processing strategies while participants watched the subtitled excerpt. Finally, as explained in the next subsection, we appraised the influence of subtitle line segmentation quality on attention allocation and performance, investigating another issue which has both theoretical and applied implications.

Subtitle Line Segmentation

A number of studies in the field of audiovisual translation focused on various aspects pertaining both to the translational and structural criteria of subtitling. As for the former, various ways to find the closest formal and semantic target language equivalents whilst reducing the source text and converting it from spoken to written have been extensively studied and finalized (Chiaro, Heiss, & Bucaria, 2008; Díaz Cintas & Remael, 2007; Gottlieb, 1992, 1994, 1998; Ivarsson & Carroll, 1998). As a consequence, well established translational strategies exist and are followed by most professional subtitlers. Recommended structural standards (e.g., guidelines for presentation time, font type, etc., see Díaz Cintas, 2003) are instead not clearly supported by appropriate empirical evidence. As in the case of other structural aspects, subtitle line segmentation has not been examined empirically. Although reference to this topic is found in almost any article dealing with subtitles, the subject has always been tackled superficially and in a passing way, with scholars encouraging, mainly on an intuitive or aesthetic basis, a coherent line parsing that does not disrupt constituency (Díaz Cintas, 2001, 2003; d’Ydewalle, Praet, et al., 1991; Ivarsson & Carroll, 1998; Karamitroglou, 1998;
Kruger, 2008; Rundle, 2000). Precise norms on this, however, do not exist, and only empirical findings will enable the appraisal of the real value of current practice.

Investigating the influence of subtitle segmentation quality has also theoretical merit. Psycholinguistic literature on reading mainly supports the idea that, since normal reading is a sequential and holistic process which occurs in chunks (Coltheart, 1987; d’Ydewalle & Van Rensbergen, 1989; Ferreira & Henderson, 1995; Rayner, 1998; Rayner & Morris, 1990; Rayner & Pollatsek, 1987, 1989), subtitle lines (or text lines, in general) should end at natural linguistic breaks, that is, ideally, at clause or phrase boundaries. This view may lead to the claim that text processing will be more difficult when segmentation does not match natural linguistic breaks because a greater integrative effort will be required (Perego, 2008a, 2008b). If ill-segmented text lines disrupt constituency and, therefore, the natural reading flow, concentration may be lost and comprehension may fail. On the other hand, one might claim that suboptimal line segmentation will not significantly disrupt processing because reading is a well-learned and partly automatized skill that will not be seriously affected by a modest increase in text integration demands. One of our aims in this article is to compare experimentally these two opposite stances and, therefore, to appraise whether ill-segmented subtitles can actually be accounted for as a real burden on cognitive processing.

AN EYE-TRACKING EXPERIMENT ON SUBTITLED FILM PROCESSING

As previously explained, we carried out an eye-tracking experiment with three main aims: a) to test experimentally the hypothesis of the cognitive effectiveness of subtitle processing by analyzing the relationship between word recognition and scene recognition; b) to test experimentally the influence of subtitling segmentation quality on cognitive processing and recognition capacity; and c) to provide new empirical evidence on the cognitive processing of subtitled programs by considering both performance measures and eye-movement analysis. Before running the study, we formulated specific hypotheses by relying on previous research and on the arguments already presented in the introduction of this article.

Processing Effectiveness

Two alternative hypotheses on processing effectiveness were contrasted. The first hypothesis predicts a tradeoff between text processing and image processing. This hypothesis stems from the ideas that watching a subtitled program is a task that requires a significant degree of attentional resources
(e.g., Grillo & Kawin, 1981), and that attention needs to be flexibly allocated on parallel information sources during this task. Thus, when attention is more focused on the subtitles, image processing will be less effective (and vice versa). This hypothesis is generally consistent with attentional theories that postulate early selection of information channels (Broadbent, 1958; Treisman, 1968; see also d’Ydewalle & De Bruycker, 2007). It is also consistent with the literature on television-message processing that postulates a potential overload in the processing capabilities of viewers, whose ability to process information is limited, especially in the presence of nonisomorphic sources conveying information that requires the division of attention over different stimuli (e.g., Bergen et al., 2005; Drew & Grimes, 1987; Lang, 2000, 2001). To summarize, according to the first hypothesis, a negative correlation between text recognition and scene recognition measures should be obtained.

According to an alternative view, processing subtitles while watching a program should not be particularly taxing from a cognitive point of view, because reading and visual processing are highly efficient and partly automatized cognitive activities (e.g., LaBerge & Samuels, 1974; Lang, 2000, 2001; Lang et al., 1999; Logan, 1997; Zhou, 2004). Moreover, according to some scholars (cf., d’Ydewalle & Gielen, 1992), their coordination within a coherent and redundant informational context does not appear to pose major problems and seems to be based on smart strategies for attention allocation and information integration. If this is the case, no tradeoff between text processing and image processing should be observed, and participants should reach relatively good levels of performance in both subtitle and scene recognition tasks. Additionally, individual variations might emerge, with participants who are more proficient in subtitle processing being also more proficient in image processing, due to individual differences in attentional resources or executive control (e.g., Del Missier, Mäntylä, & Bruine de Bruin, 2010; Friedman et al., 2007; Miyake et al., 2000).

Subtitle Quality

Two hypotheses on the effects of subtitle segmentation quality were contrasted. According to the first one, ill-segmented subtitles could significantly hinder information processing (Perego, 2008a, 2008b), slowing down reading, and causing a significant decrease of performance in text recognition (vs. well-segmented subtitles). An alternative hypothesis is that participants should not be seriously hindered by the quality of subtitle segmentation, given that reading subtitles is a partly automatic activity that requires low effort (e.g., d’Ydewalle & De Bruycker, 2007; d’Ydewalle & Gielen, 1992). Thus, no marked differences in performance should be observed, even if some differences in processing might exist.
Attentional Processing and Eye Movements

On the basis of previous research, we expected to observe more fixations on subtitles (due to the reading process) but longer fixations on the images (fixations while reading subtitles are usually rather short; e.g., d’Ydewalle & De Bruycker, 2007). Attentional shifts from text to images (and vice versa) were also expected, given that participants normally try to process and integrate visual scene information and information coming from the written text (d’Ydewalle & Van Rensbergen, 1989).

METHOD

Participants

Forty-one undergraduates and postgraduates from the universities of Trieste and Pavia (23 female, 18 male, mean age = 25, SD = 7) took part in the experiment. They were Italian native speakers with normal or corrected-to-normal (contact lenses or glasses) vision. All participants reported not being habitual viewers of subtitled films. No participant had any knowledge of the original language of the film fragment used in the experiment (Hungarian). Participants were randomly assigned to two groups (n1 = 20, n2 = 21) in order to counterbalance the within-subjects presentation of well-segmented and ill-segmented subtitles (as explained below). They were informed that the study was on subtitled-film watching. Due to time constraints, eye-movement recordings were carried out on a randomly selected subset of participants (n = 16). However, the size of the subsample of participants selected for eye-movement recording was greater than the sample sizes observed in the majority of previous studies on subtitle processing that employed eye-movement analysis.

Design

We manipulated within subjects the quality of subtitle line segmentation (high-quality, i.e., well-segmented vs. low-quality, i.e., ill-segmented), counterbalancing between subjects the order of administration of the two subtitling conditions. Thus, two different types of subtitling segmentation quality (well- vs. ill-segmented) were presented on the same video to the participants (see also the description of the video), counterbalancing their order. The main dependent variables were measures of performance on word recognition and scene recognition. Manipulation checks for subtitle reading and gist comprehension were also administered (see the following subsections for detailed descriptions).
Subtitled Video

A 15-minute video excerpt was used in the experiment. The video was taken from a Hungarian drama (*Szerelmesfilm—Love film*, I. Szabó, 1970) and had Italian subtitles. The subtitles were created using PsychInQuery (Ludwig, 2002), a well-known reliable freeware for subtitle editing. The main established criteria for line length, characters per line, synchronisation and presentation time were followed (Díaz Cintas, 2001, 2003; Ivarsson & Carroll, 1998). The overall number of subtitles created was 171; 81 sentences occupied one line of text and 90 occupied two lines. The two-line subtitles relevant for the line segmentation analysis (target subtitles) were 28. None of these subtitles was presented in the first one and a half minutes or in the last one and a half minutes of the film fragment. Each target subtitle contained various types of noun phrases (NPs) treated differently in terms of constituency (see below). We decided to choose as our target focus nonidiomatic NP structures. In particular, for the line segmentation manipulation, we considered the following NP types:

1. Noun + Adjective (e.g., *capelli biondi* “blond hair”);
2. Noun + Prepositional Phrase (PP; e.g., *colpi di fucile* “shots of a rifle”);
3. Adjective + Noun (e.g., *preciso ricordo* “precise memory”);
4. Determiner + Noun (e.g., *quel giorno* “that day”).

The target subtitles were segmented according to a simple rule that enabled us to generate two sets of 14 items each varying in their segmentation quality: In the case of all the NP constructions, well-segmented subtitles were never allowed to break into two different lines, thus, disrupting constituency, while ill-segmented subtitles systematically broke the two-line blocks disrupting constituency (see Figure 1). In particular, six Noun + Adjective target items were well-segmented (e.g., *capelli biondi* — “blond hair” — on the same line, see Figure 1) and six were ill-segmented (e.g., *capelli* and *biondi* — “blond” and “hair” — on different lines). In a similar vein, four Noun + PP target items, two Adjective + Noun target items, and two Determiner + Noun target items were well-segmented, while a corresponding number of target items of the same NP types were ill-segmented.

As for the content of the video, there was no conflicting information between video and written text (subtitles), and the only recognizable audio information came from the paralinguistic and extralinguistic elements of spoken communication, as the verbal strand had been deliberately chosen to be unknown to all participants. In the movie fragment, several stories were told with close-up of the speakers, and information was mainly conveyed through the subtitles. The topic (bad memories of socialist Hungary) was unfamiliar to most young Italian viewers but, according to a pre-test, not too easy or too complex to understand. The information was presented
in a structurally simple way: good luminance level, plain and undisturbing sound effects, low number of cuts, rather slow presentation and speech rates (cf., Fox et al., 2007; Lang, 1995, 2000).

Manipulation Checks

Two manipulation checks for subtitle reading and general comprehension were administered through questionnaires, with the aim of verifying whether participants actually read subtitles and obtained a satisfying understanding of the gist of the movie. These are two preliminary conditions that need to be satisfied before considering more specific aspects of processing performance such as word recognition and scene recognition.

QUESTIONNAIRE ON SUBTITLE READING

This short instrument comprised a series of seven questions aiming at appraising whether participants paid attention to the subtitles and to collect their opinion on the usefulness of subtitles for their understanding. Five self-report questions on a 7-point scale asked (1) how frequently the participants used subtitles to help their understanding of the film (from never to always); (2) how useful were the subtitles for their understanding (from disturbed a lot to helped a lot); (3) how difficult they found reading subtitles (from very difficult to very easy); (4) how difficult they found following the film (from very difficult to very easy); and (5) how they judged their understanding of the film (from very bad to very good). Moreover, we checked whether participants were able to remember the alignment and the color of the subtitles, using two multiple-choice factual questions.
The gist comprehension questionnaire was a multiple-choice questionnaire with 12 questions about the content of the video. The aim of the questionnaire was to appraise whether participants understood the main conceptual aspects of the film fragment and of the dialogue. For each question, participants were asked to complete a statement by selecting a response from a list of three items including a correct response and two distracters (e.g., “In the video, two young friends meet some other people. Who are these people?” potential answers: “Old Hungarian friends of the young woman,” “Hungarian immigrants who got to know each other in France,” “Hungarian relatives of the young woman”). A gist comprehension measure was computed as the proportion of correct responses. The split-half reliability of the questionnaire was .55 (computed on questions about the first part vs. second part of the video).

Dependent Variables
The two main dependent measures of the study were subtitle recognition and scene recognition. The first measure was obtained through a multiple-choice questionnaire, while the second one was acquired from a computerized recognition test.

Subtitle recognition questionnaire
This questionnaire appraised the ability to recognize specific words or short phrases presented in the subtitles (see also Gielen, 1988; Koolstra & Beentjes, 1999) through multiple-choice questions about the exact wording of the target subtitles (i.e., the subtitles whose segmentation was experimentally manipulated). The questionnaire included 28 items, divided into 14 items for each part of the video fragment. Participants had to complete a series of statements by selecting the correct response among three possible options. Distracters were semantically very close to the correct response (e.g., “In the video, Agnes says that her job made her travel a lot in various Hungarian...”; potential answers: “villages,” “cities,” “towns”). A word recognition measure was computed as the proportion of correct responses. The split-half reliability of the questionnaire was .61 (computed on questions about the first part vs. second part of the video).

Scene recognition test
The scene recognition test adopted a standard recognition procedure. Sixty freeze-frames were presented in random order on the computer screen. Participants had to decide whether each freeze-frame presented on the screen was part of the video or not by pressing two keys. Response was self-paced. Administration of the task and response recording were handled by the
FIGURE 2 Example of freeze-frames used in the scene recognition test. The panel on the left shows a frame which was presented in the video (target), while the panel on the right shows a frame not presented (foil).

e-prime software (Schneider, Eschmann, & Zuccolotto, 2002). Half of the freeze-frames used in the recognition test (15 from each of the two parts of the video) were taken from the excerpt that was actually shown during the experiment, while the remaining 30, which participants had not been exposed to, were very similar and were captured from the same film. The foil freeze-frames were selected by two independent judges, who tried to maximize the similarity between targets and foils (an example is presented in Figure 2). From the results of the recognition test we computed a proportion of correct recognition (hits + correct rejections/number of stimuli).

Procedure

Once in the lab, each participant was tested individually. Participants were given instructions and a general introduction: “You’ll be exposed to an excerpt of a Hungarian film with Italian subtitles. Your task is to watch the video and try to understand it. The film is about two young friends who, after a long time spent apart, meet in France where the young woman lives.”

Participants were seated in a comfortable chair with eyes at a distance of approximately 60 cm from the screen and they were exposed to the 15-minute subtitled video excerpt. Headphones were used to minimize external noise and distraction. As previously explained, eye movements were collected for a randomly selected subset of participants. These participants watched the subtitled video after a short calibration procedure lasting less than half a minute (see below). No other change was applied to the basic procedure.

Soon after viewing the 15-minute subtitled video, participants were asked to fill out the three questionnaires in the following order: questionnaire on subtitle reading, gist comprehension questionnaire, and word recognition
questionnaire. After completing the last questionnaire, they underwent the computerized scene recognition task. The overall time to complete the experiment was approximately 40 minutes.

**Eye-Movement Recording**

Participants’ eye movements were registered using a Tobii 1750 eye-tracking system (Tobii Technology AB, Stockholm, Sweden), which integrates all its components (camera, infrared lighting, etc.) into a 17-inch monitor. With an accuracy of 0.5 degrees and a relatively high freedom of movement, the system is a satisfying eye-movement recording solution for natural-use settings, assuring head-motion compensation and very low drift effects. It uses binocular eye tracking (i.e., for each gaze data item, data is collected simultaneously for both eyes) and the sampling rate of the device is 50 Hz (i.e., gaze data are acquired 50 times a second). In our experiment, the minimum fixation duration was set to 100 ms, and the fixation radius to 30 pixels (i.e., all the consecutive gaze points pertaining to the same fixation stayed within a circle having a radius of 30 pixels). Following a conventional procedure, the calibration in our study simply consisted in following a moving circle on the screen for 20 seconds.

**RESULTS**

In this section, we will first present the results of manipulation checks and then focus on performance measures and their relationships. After that, we will summarize the results of the eye-movement analysis on the target subtitles and relate the two sets of analyses.

**Manipulation Checks**

Participants reported having used subtitles *always* or *often* ($M = 6.6$, $SD = 1.02$, 95%CI: 6.3–6.9), and they stated that the availability of subtitles *helped* or *helped a lot* their understanding of the film ($M = 6.6$, $SD = 0.7$, 95%CI: 6.3–6.8). Participants’ memory of the visual appearance of subtitles was generally in line with their self reports; 95% of the sample correctly remembered the alignment of subtitles and 68% of the sample correctly remembered their color. These findings, together with eye-movement results, showed that participants read subtitles and relied on them to understand the film.

Participants also stated that reading subtitles was *rather easy* or *easy* ($M = 6.00$, $SD = 0.93$, 95%CI: 5.7–6.3). The last experimental condition experienced by participants did not affect this judgment (high-quality subtitles vs. low-quality subtitles: $t(39) = -0.169$, *ns*). Participants deemed the film as *rather easy* to follow ($M = 5.12$, $SD = 1.03$, 95%CI: 4.8–5.4) and they
judged their understanding as *rather good* \( (M = 4.95, SD = 1.07, 95\% \text{CI: } 4.6-5.3) \). Thus participants reported having experienced no major problems in reading subtitles and following the film. Generally speaking, the task was subjectively perceived as rather easy.

Participants showed good accuracy in their understanding of the gist of the film, both in its first part and in its second part (proportion of correct responses: first part = .75, second part = .91). However, a *t* test highlighted a significantly better understanding of the second part of the film \( (t(40) = -5.87, p < .0001) \). This can be related to the fact that, especially for an unaccustomed audience, the initial segment of a subtitled film can be less familiar or generally harder to understand (vs. the following parts), or to the actual decrease in complexity in the second part of the video used, where fewer conceptual and event changes occurred.

**Performance Measures: Subtitle Word Recognition and Scene Recognition**

**Subtitle Word Recognition**

For the participants it was harder to recognize the actual words presented in the subtitles than to answer general questions about the video. This can be easily accounted for in terms of the greater difficulty of the word recognition test as compared to the gist comprehension questionnaire. However, the proportion of correct responses showed a rather good word recognition performance (first part: .61, second part: .70). A 2 \( \times \) 2 mixed analysis of variation (ANOVA) with the test (first part, second part) and the respective position of high-quality and low-quality subtitles as factors (high-quality first, low-quality first) did not show any effect that could be attributed to the quality of subtitling. Only the main effect of the test was significant \( (F(1, 39) = 11.17, MSE = 0.013, p < .01) \), while the position of high-quality and low-quality subtitles and the position \( \times \) test interaction did not yield significant results. To conclude, we did not observe any significant effect of the quality of subtitling in the first or in the second test (first test: well-segmented .61 correct, ill-segmented .61 correct; second test: well-segmented .72 correct, ill-segmented .68 correct). As observed in the case of gist comprehension, word recognition was significantly more difficult in the first part of the film.

**Scene Recognition Test**

The proportion of correct recognition of the scenes was very high for both parts of the film (.90 and .87). Overall, participants’ performance was surprisingly good, given that they had to read subtitles in order to understand the film and that they could not rely on the Hungarian soundtrack. A mixed ANOVA did not highlight any effect of the quality of subtitling on scene
recognition. As observed for word recognition, only the main effect of the test was significant ($F(1,39) = 6.75, MSE = 0.003, p < .05$), while the position of high-quality and low-quality subtitles and the position × test interaction did not yield significant results. Thus, we did not find any significant effect of the quality of subtitling in the first or in the second test (first test: well-segmented .90 correct, ill-segmented .90 correct; second test: well-segmented .88 correct, ill-segmented .86 correct).

**Correlations between recognition measures, and other correlations**

We analyzed the correlation between word and scene recognition in order to test the hypothesis of a tradeoff between subtitle processing and image processing. The overall correlation between these two measures was not significant ($r = .19, ns, n = 41$). Similarly, the correlation was not significant in the first part of the film ($r = -.04, ns, n = 41$). However, we detected a significant positive correlation between word recognition and scene recognition in the second part of the film ($r = .32, p < .05, n = 41$). These findings are at odds with the hypothesis of a tradeoff between subtitle processing and scene processing, and consistent with the idea that watching a subtitled movie is cognitively effective.

The correlation between gist comprehension and word recognition was positive and marginally significant overall ($r = .28, p = .07, n = 41$), in the first part of the movie ($r = .28, p = .07, n = 41$), and in the second part of the movie ($r = .30, p = .05, n = 41$), while we did not find significant correlations between scene recognition and gist comprehension (overall: $r = .13, ns, n = 41$, first part: $r = -.05, ns, n = 41$, second part: $r = .10, ns, n = 41$). Another interesting result is that we did not find significant positive correlations between the self-reported comprehension evaluation and objective indices of gist comprehension ($r = -.17, ns, n = 41$) and word recognition ($r = -.09, ns, n = 41$). From these latter findings, it seems that participants are rather poor judges of their capacity to understand a subtitled film.

**Eye-Movement Analysis**

**Fixations on subtitles and on visual scenes**

Not all subtitles were used in the analysis. The eye-movement data used for it exclusively pertained to the 28 two-line manipulated subtitles containing well- vs. ill-segmented noun phrase constructions. These were the subtitles relevant for the line segmentation experiment. Data regarding these subtitles were recorded by a software (*Eye Scanpath Explorer*) specifically designed to simplify the analysis of subtitled material by grouping eye data into two
different sets: one for the area below a threshold line (subtitle region), and one for the area above it (main film zone or upper area). The default position of the threshold line was settled at 675 pixels from the top of the 1280 × 1024 screen (note that since the screen resolution was 1280 × 1024 but the video was a standard 720 × 576 DVD presented in a centred position, the actual position of the threshold line from the top of the screen on the 576-pixel high video was 451). Fixations occurring above the threshold line were considered as fixations on the main film zone, whereas fixations occurring below it were considered fixations on the subtitles.

A greater number of fixations was observed in the subtitled area of the screen versus the upper area, \( t(15) = -8.53, p < .0001 \). The number of fixations in the subtitled areas of interest was indeed more than three times greater than the number of fixations in the upper area (subtitling: \( M = 172.81 \), scene: \( M = 50.75 \)). However, the mean fixation duration in the subtitled area was significantly shorter than the mean fixation duration on the visual scene (subtitles: \( M = 221 \) ms, scene: \( M = 422 \) ms, \( t(15) = 3.57, p < .01 \)). This pattern of differences was observed both when subtitles were ill-segmented and when they were well-segmented (see the specific subsection below). Overall, participants spent 67% of the fixation time by examining the subtitled area, while the remaining fixation time was spent on the upper area of the screen. Although a greater number of fixations were observed on subtitles, fixations on the visuals were more prolonged. A qualitative analysis of fixations on a target position showed that shorter fixations on subtitles spanned over the whole written text, while longer fixations on the visual area tended to focus on specific, visually salient and attention-capturing parts of the scene (in particular, the actors’ faces; see Figure 3).

**FIGURE 3** Gaze plot showing five short fixations spanning over the subtitled area (and ranging from 140 to 279 ms) and a single prolonged fixation (1356 ms) on the speaking human face.
**PATH LENGTH AND VISUAL SHIFTS**

The path length is defined as the sum of the overall lengths, in pixels, of the segments that join each fixation point during the time interval in which each subtitle was presented. The number of visual shifts is instead the number of transitions between the two target areas of the screen (subtitled zone and main film zone) that were recorded during the time interval in which each subtitle was presented. A transition was detected any time two consecutive fixations were recorded in different areas of the screen (e.g., above and below the threshold line and vice versa).

The number of fixations was found to be strongly related to the path length \( (r = .93, p < .001, n = 16) \). However, the correlation between the path length and the number of fixations was significant only for the fixations in the subtitled region \( (r = .90, p < .001, n = 16) \) but not for the fixations in the main film zone \( (r = .36, ns, n = 16) \). This is an indication that eye movements were mainly devoted to subtitle reading and only in a minor part to the analysis of the visual scene (i.e., participants made more fixations on the subtitles than on the upper part of the screen). A positive relation was also detected between the number of fixations and the number of visual shifts \( (r = .64, p < .01, n = 16) \). More specifically, the number of shifts between subtitles and visuals was related to the number of fixations in the upper part of the screen \( (r = .69, p < .01, n = 16) \). Participants who made more fixations on the visual scene were obviously more likely to make a greater number of shifts from subtitles to visual information.

**DIFFERENCES BETWEEN WELL-SEGMENTED AND ILL-SEGMENTED SUBTITLES**

No significant difference in the number of fixations was found between high-quality and low-quality subtitle conditions (high-quality: \( M = 111.50 \), low-quality: \( M = 112.06 \), \( t(15) = -0.09, ns \)). Similarly, no differences between these conditions were observed in the number of fixations on the visuals (high-quality: \( M = 26.25 \), low-quality: \( M = 24.50 \), \( t(15) = 0.74, ns \)) or within the subtitled area (high-quality: \( M = 85.25 \), low-quality: \( M = 87.56 \), \( t(15) = -0.40, ns \)). The proportion of the fixation time devoted to the subtitled area did not differ significantly between conditions either (high-quality: 0.69, low-quality: 0.69, \( t(15) = -2, ns \)). No difference in the mean fixation duration was found between the well-segmented and the ill-segmented conditions (high-quality: \( M = 259 \) ms, low-quality: \( M = 261 \) ms, \( t(15) = -0.268, ns \)). In the same vein, no difference was observed in the mean fixation duration in the visual areas (high-quality: \( M = 433 \) ms, low-quality: \( M = 433 \) ms, \( t(15) = 0.001, ns \)). However, a significant difference, albeit very small, was detected in the mean fixation duration within the subtitled area. In particular, fixations on ill-segmented subtitles were slightly longer than fixations on well-segmented subtitles (high-quality: \( M = 216 \) ms, low-quality: \( M = 228 \) ms, \( t(15) = -2.27, p < .05 \)). However, the ratio between the
mean fixation duration above and within the subtitled area did not differ across conditions (high-quality: 1.97, low-quality: 1.87, \( t(15) = 0.56, \text{ns} \)). No significant differences between conditions were observed in the path length (high-quality: \( M = 12599 \), low-quality: \( M = 12529 \), \( t(15) = 0.09, \text{ns} \)) or in the number of visual shifts between the subtitled area and the upper region of the screen (high-quality: \( M = 19.12 \), low-quality: \( M = 16.87 \), \( t(15) = 1.65, \text{ns} \)). Summarizing, no differences were observed between the high-quality and the low-quality subtitles, with the exception of a very small difference in the mean fixation duration within the subtitled region.

**Relationships between eye movements and performance measures**

No significant (or even marginally significant) relations were found between eye-movement measures (number of fixations on the different areas of the screen, mean duration of these fixations, path length, visual shifts between subtitled areas and visual areas) and behavioral measures on subtitle word recognition and scene recognition. Additionally, no eye-movement measure was found to be related to recognition measures after the disaggregation of the data in well-segmented and ill-segmented subtitles.

**Differences between the first and the second part of the film excerpt**

There were some differences between the first and second part of the video excerpt in terms of eye movements. Indeed, more fixations were observed in the second part of the film (first part: \( M = 104.87 \), second part: \( M = 118.69 \), \( t(15) = -3.13, p < .01 \)). However, significantly more fixation were recorded only in the subtitled area (first part: \( M = 78.00 \), second part: \( M = 94.81 \), \( t(15) = -4.44, p < .001 \)), while there was a nonsignificant decrease in the number of fixations on the visuals (first part: \( M = 26.87 \), second part: \( M = 23.87 \)). The mean fixation time on the subtitles did not change significantly (first part: \( M = 219 \text{ ms} \), second part: \( M = 224 \text{ ms} \)), while the mean fixation time on the visuals was significantly longer in the second part of the film (first part: \( M = 392 \text{ ms} \), second part: \( M = 475 \text{ ms} \), \( t(15) = -2.19, p < .05 \)). Finally, the path length was significantly longer in the second part of the video (first part: \( M = 11661 \), second part: \( M = 13469 \), \( t(15) = -2.91, p < .05 \)), while the number of visual shifts between the subtitled area and the upper region of the screen did not differ significantly (first part: \( M = 17.87 \), second part: \( M = 18.12 \)).

These differences seem to indicate that participants changed some aspects of their processing in the second part of the film excerpt, devoting increased attention to the subtitles, but also increasing, at the same time, the mean fixation time on the visuals. However, this change did not strongly affect the participants' basic processing strategies. In fact, as far as the second part of the video is concerned, the eye-movement analysis highlighted still
much more fixations on the subtitles than on the visuals, still longer fixations on the visuals, a constant number of visual shifts, the same tendency to focus on the subtitle text and on salient aspects of the visuals (such as the actors’ faces). Apparently, participants fine-tuned or slightly adapted (but did not change sharply) their processing strategies in the second part of the film. This could have happened as a consequence of learning processes or as an adaptation to specific movie contents or features. According to our view, this pattern of findings seems more consistent with the idea that participants improved or adapted their processing strategies in the second part of the movie than with the idea that the second part of the movie was simply easier to understand. If the latter possibility were true, we should have not observed increased fixations devoted to the subtitle area and in the path length measure.

DISCUSSION

The experiment described in this article was carried out with three aims. The first aim was to test experimentally the hypothesis of the cognitive effectiveness of subtitled program processing (e.g., d’Ydewalle & Gielen, 1992). The results we obtained are generally consistent with this hypothesis, showing good levels of performance both in subtitle recognition and scene recognition and, more importantly, highlighting the absence of a tradeoff between subtitle recognition and scene recognition. A striking indication of the effectiveness of participants’ cognitive processing was the remarkable performance in the scene recognition task observed, even if less than 40% of the fixation time was devoted to visual scenes. Good performance in freeze-frames recognition appears to be in line with previous research on audio-video processing (e.g., Drew & Grimes, 1987). As for the word recognition performance, it was good but not striking, possibly because our recognition test required a very fine discrimination between semantically similar responses (unlike previous studies, where recognition tests pertained to rougher discriminations of whole subtitles, e.g., Gielen, 1988; Koolstra et al., 1999). The phenomenon of picture superiority could also contribute to explain the observed difference between scene recognition and subtitle word recognition (Lang, 2001; see also Grimes, 1991; Paivio, 1971, 1986; Zhou, 2004).

More importantly, we did not observe any sign of a tradeoff between word recognition performance and scene recognition performance, but a significant positive correlation in the second part of the film. This positive correlation might be explained by assuming that individuals with greater attentional capacity or executive control can process more comfortably both subtitles and scene information, and that these individual differences can be captured when the difficulty of the task allows their detection or when partic-
Participants have sufficient practice with the task. While this potential explanation deserves to be further explored, our findings clearly show that participants did not tradeoff subtitle processing and scene processing. To summarize, our global pattern of findings is consistent with the idea that processing subtitled programs is cognitively effective.

The second aim of our study was to test experimentally the influence of subtitle segmentation quality on cognitive processing and recognition, starting from psycholinguistic considerations on this issue, and aware that structural and typographical factors may come into play in the reading process, as demonstrated by some studies (e.g., Beymer, Russell, & Orton, 2007; Chaparro, Shaikh, & Baker, 2005). No indication of differences in processing or performance related to subtitling segmentation quality was found over a large set of measures, with the single exception of a very slight difference in subtitle fixation mean time. This enables us to conclude that subtitle segmentation quality did not have a significant impact in our study. Participants appeared to process well-segmented and ill-segmented subtitles basically in the same way and with the same outcomes. We, therefore, conclude that psycholinguistic concerns about subtitle line segmentation are probably overstated.

The final objective of our study was to provide new empirical evidence on the cognitive processing of subtitled films by considering both performance measures and eye-movement analyses. In this regard, we obtained some results consistent with previous research, but we also acquired novel findings and some results that are not completely in line with previous studies. Consistently with previous research, a significant number of participants’ fixations (and fixation time) was devoted to the subtitled area. We also observed the expected visual shifts from subtitles to visual scenes (and vice versa, see de Linde & Kay, 1999b; d’Ydewalle & Van Rensbergen, 1989), which enable participants to process both sources of information.

Novel results were instead obtained on the effectiveness of subtitle processing (in particular on scene recognition capacity) and on the absence of a tradeoff between image processing and subtitle processing. Furthermore, we observed that more fixation time was devoted to subtitle reading than to the visual analysis of the film scenes, but fixations on visuals were longer than fixations on subtitles. This suggests that participants, while watching the film, read the subtitles in order to understand the story, but they did not perform an extensive exploration of the overall visual scene, focusing instead their attention only on the most informative or visually salient elements (Lautenbacher, 2009). Although some gist information can be acquired from a brief glance at a scene (Schyns & Oliva, 1994), selective visual processing of scenes or pictures is needed to elaborate visual details (Liversedge & Findlay, 2000). We observed that our participants tended to focus on faces, and this agrees with the well-documented natural tendency to notice and focus on faces when looking at a scene or picture (e.g., Brown, Huey, &
Findlay, 1997; Theeuwes & Van der Stigchel, 2006; for similar finding in primates, see Kano & Tomonaga, 2009, 2010; Tomonaga & Imura, 2009). Face is a natural attention-grabbing visual cue, possibly because it conveys essential information for understanding interaction and communication (including emotional signs).

Another finding that deserves to be commented is the absence of significant relationships between individual differences in eye movements and recognition performance. This can be explained by taking into account the rather uniform eye movement behavior exhibited by participants during the task (the great majority of them made more fixations and globally spent more time on the subtitles, and made fewer and longer fixations on the visual scene). This behavior seems to be associated with good levels of recognition and understanding in our experiment.

Finally, results not completely consistent with d’Ydewalle and De Bruycker (2007) were found for saccade mean duration on two-line subtitles (221 ms in our study vs. 179 ms in d’Ydewalle and De Bruycker) and for the proportion of time spent on the subtitles (67% in our study vs. 37%—adult sample, two-line subtitles—in d’Ydewalle and De Bruycker, 2007, p. 199). These differences can be explained by the different type of video used in the two experiments. While in our video following the dialogue rather than the visuals was essential for an appropriate understanding, d’Ydewalle and De Bruycker’s animation possibly allowed participants to allocate less attention on subtitles. In any case, it is interesting to note that our mean fixation time on the subtitles was in line with the values reported by d’Ydewalle and De Bruycker as standard for ‘normal reading’ (200–250 ms).

To summarize, participants seem to deploy efficient and selective attention allocation strategies which enable them to reach a good understanding and a good recognition of presented information. If we consider our findings in the light of previous research, we can also hypothesize that participants modulate their processing strategies according to the type of subtitled program they are watching (see also d’Ydewalle & Gielen, 1992). However, further research employing different film types is needed to understand if this idea has empirical support.

The results of the present study are particularly interesting if we consider that our participants belong to a classical dubbing country (Italy), and they reported to have a limited experience with subtitled programs. Given that our findings generally support the idea that the processing of a subtitled film is cognitively effective, we can expect even more encouraging results in samples of participants belonging to subtitling countries (e.g., Scandinavian countries, the Netherlands, and Greece: Luyken et al., 1991; see also Brondeel, 1994; d’Ydewalle & Gielen, 1992). On the other hand, we expect that our basic findings will be replicated in other nonsubtitling countries (e.g., the United States and the United Kingdom). Furthermore, the young age and relatively high education level of our participants suggest the need
for understanding whether the present findings can be also generalized to different age groups (e.g., children and older adults) and to respondents with a lower level of education.

Theoretical, Methodological, and Applied Implications

From the theoretical point of view, our findings raise both general and specific implications. Specific implications concern the theoretical account of subtitle processing. As we have just specified, our results support the idea that subtitled material processing is cognitively effective and that processing strategies are possibly adaptively modulated to comply with specific task demands. Moreover, the results also showed the cognitive robustness of subtitle processing, which is not affected by a major structural change such as subtitle line segmentation. Future research could explore three promising directions. First, it would be worth examining systematically how specific features of subtitled programs affect processing strategies (e.g., the relative importance of dialogue and visuals, the complexity of the plot, the degree of text-image redundancy). Second, the boundaries of the cognitive effectiveness of subtitle processing need to be explored. A tradeoff between subtitle reading and visual scene encoding might emerge if viewers are exposed to very complex programs, where information changes at a fast rate and there is little redundancy between information sources (Bergen et al., 2005; Fox et al., 2007; Drew & Grimes, 1987; Lang, 1995, 2000). The results of some studies on subtitle reading in aging (e.g., d'Ydewalle & Van Resenbergen, 1989) seem to suggest that, when cognitive resources decline, participants focus more on the less demanding source of information (visuals) and pay less attention to the more demanding source (subtitles)—which is consistent with what usually happens in TV message processing (Grimes, 1991). Third, it will be theoretically interesting and useful to explore how individual differences in executive control and attentional capacity affect the ability to process subtitled programs, and how subtitled programs are processed by older adults and by an impaired audience.

General theoretical implications of our findings concern multiple-source processing and the integration of information in communication (d'Ydewalle & Gielen, 1992; Gielen, 1988). The results of our study, in line with previous research, highlighted the individuals' ability to process, integrate and remember information coming from various sources. A possible conclusion is therefore that individuals will not usually encounter serious difficulties in multiple-source information processing. However, this conclusion needs to be qualified. First, as we pointed out, the picture can change if more complex and less redundant information has to be processed. Moreover, things may be much harder if multiple written sources of information need to be concurrently attended to (as in complex displays of news) or if more than one source delivers nonredundant verbal information in the same language.
In similar cases, greater interference has been documented by basic cognitive research (e.g., Wickens & McCarley, 2008). Finally, specific characteristics of the individuals need to be taken into account (age, education, attentional abilities, processing speed, etc.). Further research along these directions is thus needed to qualify the existing findings and to provide a more complete picture of multiple-source information processing in communication.

From the methodological point of view, we suggest that the experimental study of subtitling and multiple-source information processing should take into account a variety of measures related to participants’ general understanding, word recognition, and capacity to recognize scenes and process visual information. In fact, given that these different measures do not appear to be strongly correlated, they are all needed to gain a representative picture of participants’ performance. Moreover, the analysis of the correlation between text-related and image-related measures can be informative about the presence of text and scene processing tradeoffs that might be observed when the program is very difficult to understand, when different sources of information interfere, or in individuals with reduced attentional resources. The present study also suggests that it is not wise to rely on participants’ self-reported evaluation of their comprehension, given that self-reports do not always correlate with objective measures. Finally, although eye-movement analysis can be useful to obtain indications about the direction of visual attention and on processing strategies, we think that it must always be complemented with a set of performance measures in order to provide a fully informative picture.

In conclusion, it is worth mentioning some applied implications of our findings. First, our study confirms that subtitled programs are usually rather easy to follow. This is good news for subtitle professionals, who can rely on a comfortable solution when dubbing is too costly. Moreover, we observed that the cognitive processing of subtitles is not affected by the violation of an important structural guideline. While we encourage professionals to continue adopting regular line breaks (for aesthetical reasons that might have an impact on the viewer’s evaluation), our findings point to the general need of empirically appraising the real effectiveness of subtitling guidelines. As our study contributed to point out (but see also Koolstra, Peeters, et al., 2002), these guidelines are often not clearly grounded in empirical studies. Therefore, we think that it will be worth promoting an applied research program providing a systematic empirical test of subtitling guidelines for different groups of potential users (e.g., young adults, older adults, and deaf people).

ACKNOWLEDGMENTS

Partial results of this study have been presented for the first time at the International Conferences, Audiovisual Translation: Multidisciplinary Approaches,
June 19–21, 2008, at the University of Montpellier 3, France, and Eye-to-IT, April 28–29, 2009, in Copenhagen, Denmark.

The authors are grateful to the Computer Vision & Multimedia Lab, University of Pavia, Italy, and thank all the volunteers who took part in the experiments, thus making it all possible. The authors also thank Consorzio Universitario di Pordenone for support for research activities, and would like to thank Professor Annie Lang and two anonymous reviewers for their careful and constructive comments on the manuscript.

NOTES

1. It is acknowledged that eye movements are only approximate indicators of the direction of visual attention: what is being paid attention to and where the eyes fixate or are directed can be dissociated, although this is usually not the case (Duchowski, 2007; Posner, 1980). The connection between gaze and visual attention is, indeed, usually close, and so is the relation between eye movements and cognitive processing activities during reading (Fisher, Karsh, Breitenbach, & Barnette, 1983; Krause, 1982; Rayner, 1998).

2. This software was created by Marco Porta at the Computer Vision & Multimedia Lab, University of Pavia, Italy.

3. For reasons of space, we do not present here a detailed account of these nonsignificant findings. However, a report on these results is available on request from the authors.

REFERENCES


In J. K. O’Regan, & A. Lévy-Schoen (Eds.), *Eye movements: From psychology to cognition* (pp. 313–321). Amsterdam, the Netherlands: Elsevier Science.


