

Emotional Responses to Aesthetically Integrated and Standard Subtitles in a Fantasy-Thriller Audiovisual Context

 Francesca Leveridge 

University of Nottingham

 Pierre-Alexis Mével 

University of Nottingham

 Myron Tsikandilakis 

University of Nottingham

Citation: Leveridge, F., Mével, P. A., & Tsikandilakis, M. (2024). Emotional responses to aesthetically integrated and standard subtitles in a fantasy-thriller audiovisual context. *Journal of Audiovisual Translation*, 7(1), 1–32. <https://doi.org/10.47476/jat.v7i1.2024.266>

Editor(s): J. Pedersen

Received: September 9, 2023

Accepted: February 3, 2024

Published: June 17, 2024


Copyright: ©2024 Leveridge, Mével and Tsikandilakis. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/). This allows for unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Abstract

Situated at the intersection of Psychology, Film Studies, Accessibility Studies and Translation Studies, this article investigates the emotional correlates of two types of subtitles (standard and aesthetically integrated) on audiences in the context of fear-eliciting clips with Russian fantasy thriller *Night Watch* (Bekmambetov, 2004). Our experiment employed a methodology combining skin conductance (SCR), heart-rate responses (HR) and self-reports (questionnaires) to account for the complex interplay between experiential, cognitive, behavioural and physiological elements that make up emotional responses. We examined the psychophysiological and self-report responses to two subtitling delivery effects – standard subtitles and aesthetically integrated subtitles – focussing specifically on fear. We used significance-testing and Bayesian analyses to compare the two subtitling deliveries. For both analyses, we found that the presentation of aesthetically integrated subtitles led to higher positively rated psychophysiological arousal and quality of viewing experience ratings compared to standard subtitles. This novel finding suggests that aesthetically integrated subtitles could play an important role in the development of new ways to provide audiovisual translation.

Key words: subtitles; fantasy-thriller; psychophysiology; emotion; skin conductance; heart rate.

 francescaleveridge@gmail.com, <https://orcid.org/0000-0002-8768-6900>

 pierre-alexis.mevel@nottingham.ac.uk, <https://orcid.org/0000-0003-3503-5302>

 tsikandilakismyron@gmail.com, <https://orcid.org/0000-0001-8829-7563>

Introduction

Over the past decade, research on audiovisual translation (AVT) has seen an incremental shift towards experimental research and approaches drawing from the fields of linguistics, psychology and cognitive science to study the reception and perception of translated audiovisual texts (Di Giovanni & Gambier, 2018). For example, audience engagement and enjoyment is now routinely monitored by means of eye-tracking devices, facial recognition software and webcams, and a smaller, but growing number of studies are relying on instruments such as electroencephalography (EEG), electrocardiograms (ECG) and skin conductance responses (SCR) to measure viewers' immersive, cognitive, affective and sensual experiences through their peripheral and central nervous system activity (see Szarkowska & Gerber-Morón, 2018; Díaz Cintas & Szarkowska, 2020, for an overview). Although more traditional methods such as questionnaires and interviews have yielded fruitful findings, they are often considered to provide sequential-ordinal data that do not involve within and between participants agreeable rating characteristics (see Boone & Staver, 2020; Bürkner & Vuorre, 2019). This means that when evaluated without physiological assessment technology (Cacioppo et al., 2007) and using the correct analyses (Wagenmakers et al., 2018a, 2018b), they could fail to capture the full array of effects that audiovisual texts have on viewers. Employing them in conjunction with this physiological assessment technology affords researchers the opportunity to delve further into the minds of audiences, assessing actions, behaviours and expressions, to ultimately arrive at a deeper understanding of the impact that different modes of AVT have on the already distinctively multimodal nature of audiovisual communication.

Audiovisual texts consist of both the visual and auditory components, with meaning arising as a result of “the relationship between image and word, the interplay of the signification systems” (Chaume, 2004, p. 24). The translation of audiovisual texts, be this through dubbing or subtitling, adds further layers of complexity to an already complex network of meaning-making resources – verbal, non-verbal, aural and visual (Gottlieb, 1994; Chaume, 2004) – that renders the interplay between them and their collective sensory import all the more intricate. In the case of subtitling, or the process of adding written text to audiovisual media to express messages being spoken in dialogues, the original aural and visual channels remain unchanged, but the semiotic balance between them shifts as text reception becomes primarily visual (Gottlieb, 1994, p. 162). The fact that subtitles call attention to the dialogues in a visual way means they compete with the other visual signals for the attention of the viewer, requiring different cognitive processes, as they must read the subtitles whilst mapping what they read onto what they see and/or indeed sometimes hear. Whilst there is substantial data – collected primarily through eye-tracking studies – showing that many viewers are able to simultaneously process the subtitles and absorb all the salient elements of the visuals without any difficulty (d’Ydewalle et al., 1991; d’Ydewalle & De Bruycker, 2007; Perego et al., 2010), it is nonetheless reasonable to assume that for others, engaging in a continuous process of reading might prove somewhat disruptive to the otherwise immersive and relaxing entertainment experience of watching audiovisual media such as films.

One of the main criticisms of subtitles is that they are distracting and interrupt the visual aspects of films. To cite just one example, Ivarsson and Carroll, the creators of the Code of Good Subtitling Practice, write that:

The disturbing subtitles crowd out the picture and ruin the composition (...) And then they divert the viewer's attention from the picture. The subtitles often flit in and out without being synchronised with the takes, in utter disregard of the film's rhythm and intention. (1998, p. 34)

That many viewers read subtitles even when they are neither needed nor understood (d'Ydewalle et al., 1991; d'Ydewalle & De Bruycker, 2007; Bisson et al., 2014) does indeed suggest that they are hard to ignore, and perhaps even more so when they appear to be at odds with the rest of the semiotic information. Clashes and inconsistencies between the images and soundtrack on the one hand, and the subtitles on the other, not only accentuates the visibility of the subtitles, but also the fact they are visibly not a part of the original audiovisual material the audience is watching. Subtitles may also be described as "additive", in that they are usually created once filming is completed; they "constitute an addition to an already autonomous product, and thus also emphasize their own non-autonomous nature" (Mével, 2017, p. 95). Subtitles that draw too much attention to themselves – be this due to technical, linguistic or visual inconsistencies – are therefore believed to break the metaphorical "contract of illusion" between the film and its audience (Pedersen, 2017), potentially lowering the viewers' enjoyment and feeling of immersion in the film's diegesis.

Subtitling guidelines such as the Code have traditionally encouraged a strategy of "invisibility" (O'Sullivan, 2011; Pedersen, 2017; Romero-Fresco, 2019), whereby the aim is to make subtitles as plain and unobtrusive as possible so that the viewer forgets they are reading them. The idea that subtitles should be invisible is somewhat of a fallacy (Foerster, 2010); subtitles are visible, and they must be, if they are to be read and used by viewers to make sense of what they are watching. Somewhat ironically, the very strategies aimed at optimising the reading experience and minimising the disturbance caused by the unavoidable presence of translation seem to produce the complete opposite effects. For instance, subtitles are typically limited to two lines and positioned at the bottom of the screen where it is assumed that less action, or less important action, will take place (Díaz Cintas & Remael, 2007, p. 82). This assumption often turns out to be false. Numerous examples can be found to show how subtitles with a conventional layout and position cover up plot-relevant images and objects (see for example Fox, 2018, p. 69; Romero-Fresco, 2019, p. 93), or fade into light backgrounds when they are presented as white text, potentially causing the viewer to miss important information that hinders their overall comprehension of the narrative, as well as lead to a loss of aesthetic value. We might also note how the use of white or yellow text, aimed at increasing legibility (Ivarsson & Carroll, 1998), is frequently at odds with the aesthetic of a film, which creates a bigger contrast between the subtitle and the image, drawing attention to their otherness, their visibility as a translation, and potentially drawing the viewer out of the fictional world.

Elsewhere, the growing popularity of "creative subtitles" (McClarty, 2012) suggests that these deeply entrenched industry practices are changing, and this popularity appears to be challenging deeply held

believes that the “best subtitles are those that the viewer does not notice” (Díaz Cintas & Remael, 2007, p. 40). Creative subtitles, also known as “dynamic subtitles” (Brown et al., 2015), “integrated titles” (Fox, 2016, 2018) and “free form subtitles” (Bassnett et al., 2022), override traditional subtitling conventions by experimenting with typeface, font size, placement on screen, for example, as well as employing animation and special effects that correspond closely to the film’s visual style as well as action. Creative subtitles are not bound to the text’s surface but are free to move within a given action sequence, providing extra layers of meaning to the narrative, themes and characterisation, sometimes even becoming a part of the story themselves. Instead of being “ancillary” (Pedersen, 2007) or an “afterthought” (Romero-Fresco, 2019), as is the case with their “conventional” counterparts, creative subtitles are deliberately, graphically integrated into an audiovisual work. They thus contribute to and become a part of its image, typographic and aesthetic identity (Romero-Fresco & Fryer, 2018, p. 13), giving the impression that the filmmaker had intended for them to be there the entire time.

Experimental research in the domain of creative subtitling has to date focussed almost exclusively on particular cognitive aspects of the viewing experience, such as information intake, reading speed and visual attention, and seemingly ignored the affective processes involved in film viewing, namely that of emotional response, which is the focus of the present article. McClarty (2012, p. 145) has detailed different ways in which creative subtitles may “respond to”, “engage with” and even “mirror the emotion of the moment” through variations to both linguistic and graphic forms, be this colours, styles or special effects. Whether the degree of integration or style of subtitles makes a difference to how audiences respond emotionally to films remains to be seen. The emotional reception of subtitled films has, more broadly, received very little attention in AVT, not least on an experimental basis, until now. This article provides an overview of a study carried out by a team of researchers at the University of Nottingham (United Kingdom) to determine the emotional correlates of subtitles that are integrated into a film’s aesthetics. Using clips from fantasy thriller *Night Watch* (Bekmambetov, 2004), the researchers used a between-subject design to measure differences in the reception of conventional versus creative subtitles in relation to a key emotion: fear.

The experiment employs a methodology combining skin conductance responses (SCR), heart-rate responses (HR) and self-reports (questionnaires) to account for the complex interplay between experiential, cognitive, behavioural and physiological elements that make up emotional responses (see Izard, 1993, for a summary). All emotional experiences are subjective, meaning the specific emotion elicited by a given event or stimulus, i.e., whether it induces feelings of pleasure or displeasure, will depend on whether it is evaluated as positive, negative, aversive, and so on, which can depend on individual circumstances, values and contexts (Mandler, 1992). The intensity with which an emotion is experienced can be measured by the strength of visceral arousal, i.e., the activation of components of the autonomic nervous system, such as the heart, and skin conductance responses (SCR; sweat gland activity) (Mandler, 1992). The subjective nature of emotional experiences means that the same emotions can be experienced to a greater or lesser degree depending on a range of outside factors, and that only the person experiencing the specific emotion is able to evaluate and rate it, although measuring heart-rate and skin conductivity allows us to

observe reactions which participants may not report, due to, for example, self-presentation and awareness biases (see Ciuk et al., 2015).¹ Whilst it is therefore likely that the impact of the cinematic techniques used to elicit fear – and how these are affected by the subtitling decisions – may vary from viewer to viewer, they should nonetheless elicit similar emotional responses across diversified audiences and produce similar emotional trajectories for viewers of the same film (Carroll, 2008).

The choice to focus on fear is due to the fact that it is one of six basic or universal emotions (Ekman, 1999) alongside anger, surprise, disgust, sadness and happiness. It is also one of the most widely explored emotions in linguistics and psychophysiology as regards the anticipated correlates of arousal it induces (Kreibig, 2010). Fear, which is generally understood to be the response to the threat of harm, be this physical or psychological, and activates impulses to freeze or flee, would be classed as an emotion with a high level of arousal and a negative valence. Like other negative emotions, it is therefore easier to induce and measure (Taylor, 1991; Uhrig et al., 2016), which makes it a useful starting point for measuring the emotional correlates of creative subtitles. The proposed methodology is therefore innovative, having yet to be applied to types of subtitling that would more traditionally (or perhaps typically, in the case of non-traditional styles such as creative subtitling) be classed under the umbrella of AVT rather than media accessibility (MA), and which stand out as an important object of study because of their vast yet unharnessed potential for providing access and for offering new underexploited pathways to storytelling.

After a brief literature review which covers the main psychophysiological studies on subtitle reception and on the affective reception of films, the article addresses some terminological and epistemological concerns around “creative subtitles”. Our focus then shifts to the experiment, starting with stimuli design, stimuli selection in the first stage of the experiment, and stimuli reception during the second stage of the experiment. We then discuss our findings.

1. Literature Review

1.1. Psychophysiological Studies on Subtitle Reception

Of the different psychophysiological methods that may be employed to measure subtitle reception, the most frequently used in the field is eye tracking. Eye tracking typically involves infrared technology optical tracking without direct contact to the eye, and can inform a researcher concerning the position, movement and fixation duration of a participant to a presented elicitor (Bergstrom & Schall, 2014). Kruger (2018), Doherty and Kruger (2018) and Szarkowska and Gerber-Morón (2019) provide useful overviews of existing studies and the range of issues eye tracking has been used to investigate, from the impact of line breaks, text segmentation and subtitle duration on visual

¹ Changes in physiology have been proven to correlate with what we feel and experience (Kreibig, 2010), and so employing affective reports alongside physiological measurements is a useful way of triangulating data.

attention and cognitive load to how the different types of subtitling affect how subtitles are processed altogether. As far as “creative” subtitles are concerned, eye tracking has also been the preferred method of experimental reception studies, as well as a general tendency to isolate the variable of subtitle placement. For example, Brooks and Armstrong (2014), Brown et al. (2015), Fox (2018) and Black (2020) used eye tracking to determine whether reducing the distance between the subtitles and characters or other areas of interest on the screen would reduce the time spent on reading. Whilst the effects of integrated subtitle position on the processing and immersion are now relatively well known, there is a noticeable lack of experimental research testing the relationship between the other distinguishing features of creative subtitles, i.e., variations on typeface, font size and use of special effects, and the cognitive (and indeed affective) processes of film viewing.

Current research is also beginning to incorporate electroencephalography (EEG) into eye tracking methodologies in order to provide a more direct measurement of immersion and cognitive load in subtitle processing. Of particular relevance here is a study conducted by Kruger et al. (2018) that used both approaches, along with self-reported measures, to investigate the impact of linguistic choices and subtitle position in the absence and presence of “standard” and “integrated” intralingual subtitles in an English fiction film viewed by L2 speakers of English. Their findings show similar levels in the viewers’ sense of immersion in the fictional reality across both types of subtitles, but a “deeper, uninterrupted processing of the content” in the latter, resulting in overall reductions in perceived complexity (2018, p. 289), as well as generally positive self-reported responses towards the integrated subtitles. To the best of our knowledge, there are no studies that use EEG to examine the emotional processing of translated audiovisual content, although it has been used to explore positive and negative emotional responses when watching films more broadly (see Petrantonakis & Hadjileontiadis, 2010, 2011; Lin et al., 2010; and Nie et al., 2011). Despite this, as Kruger et al. (2016) point out, the link between emotion and immersion has not yet been sufficiently established and remains a vast unexplored field.

1.2. Affective Reception of Films and AVT

The measurement of skin conductance (SCR) and heart rate (HR) responses are also increasingly being used to study the affective reception of films, as evidenced in studies by Dillon (2006), Codispoti et al. (2008), Fernández et al. (2012), Bos et al. (2013) and Brumbaugh et al. (2013), and more recently films with subtitles. Apart from the experiment detailed in this paper, these measures have, to date, only been applied in two other experimental studies by Iturregui-Gallardo and Matamala (2020) and Iturregui-Gallardo and Soler-Vilageliu (2021) which targeted subtitles that may be considered “unconventional”. Specifically, they focus on audio subtitling (AST), a media accessibility service that allows for people with visual or reading impairments to access written subtitles in their aural form. The former set out to examine responses to AST with a voice-over and dubbing effect and the latter AST and conventional written subtitles, and both exposed blind and partially sighted participants to clips that expressed sadness, fear or were emotionally neutral. Both studies found that there were few differences overall in the viewers’ emotional experience across each type of AST, although some

expected disparities were observed in the self-reported responses to the clips in terms of valence – negative for fear and sadness, but not for neutral – and arousal – deemed higher for fear than for sadness and neutral.

In line with the use of empirical experimentation in reception studies of translated audiovisual content more generally, the study of emotional responses to films with AVT more specifically has found significant currency in the field of media accessibility, and in particular in the domain of audio description (AD). Fryer (2013) used SCR and HR in combination with questionnaires to study viewer immersion and the elicitation of amusement, fear and sadness in clips with no AD and those featuring AD delivered by synthetic and human voices. Two later studies by Ramos (2015) and Ramos Caro (2016) using the same methods focussed on responses to the emotions of disgust, sadness and fear in clips with and without AD as well as clips using “objective” and “subjective” AD, the latter being a more intensely “descriptive” form including metaphors, inferences and subjective evaluations of the content being described. The self-reported measures in all three studies presented differentiations in response to each emotion elicited, and significant differences where the clips showing fear were concerned, concretizing the value of this emotion for experimental research and foreshadowing its potential to generate meaningful results in the present experiment. However, and as with the above-mentioned studies on emotion and AST, the results from the psychophysiological measures were mostly inconsistent, suggesting that concentrating on one emotion, rather than multiple, may produce data that are more conclusive.

2. Terminological Concerns

As is clear from the introduction to this paper, several terms have emerged in recent decades in an attempt to capture a variety of non-conventional subtitling that draws from a range of existing practices and exists in a variety of forms and in multiple contexts. For example, Díaz Cintas and Muñoz Sanchez (2006) use the term “hybrid” when discussing examples of captions that experiment with font, length and placement found in fansubs of anime and similar “amateur” subtitling practices. McClarty’s (2012) concept of “creative subtitles” was developed with intralingual captioning in mind and takes inspiration from theatre translation, which she observes is typically a collaborative practice between producer and translator and tailored to each individual production, paying attention to the “specific aesthetic qualities of the individual play that is being translated” (2012, p. 138). The more recent notion of “accessible filmmaking” (Romero-Fresco, 2019) locates “integrated” subtitles, a term coined by Fox (2016), in the context of arthouse cinema and commercial film and television, as both an intra- and interlingual practice, and one that bridges the gap between notions of subtitling as a tool that exists solely to provide access and a form of aesthetic creativity in its own right. With the distinctions between more “traditional” types of subtitling – at linguistic, practical and conceptual levels – and their respective target audiences themselves becoming ever more blurred (Szarkowska & Gerber-Morón, 2019), it is hardly any surprise that pinning down these relatively newer concepts is challenging.

There are broader debates currently circulating within AVT scholarship regarding the language being used to define subtitles that use their added communicative properties in a “non-standard” way and distinguish them from those deemed more “conventional”. The term “creative” has proven to be particularly controversial, with Pedersen (Bassnett et al., 2022, p. 10) arguing that it is “misleading to talk about creative subtitles only when they look different, e.g., when they have unusual colours, fonts or placement.” It is certainly true that “creativity is a continuum and requires intent” (Pedersen, 2022, p. 10); even those subtitles that are presented more plainly in terms of their visual style may display forms of creativity or experimentation in terms of their linguistic content. Multilingual films, and particularly those that are part-subtitled, also challenge ideas about subtitles being an “afterthought” (Romero-Fresco, 2019) as they are typically borne with the idea to make a film in multiple languages, and thus integrated into the film’s overall narrative structure (O’Sullivan, 2011; Langer Rossi, 2019; Leveridge & Mével, 2023). Of course, on a purely technical level, it is problematic to attempt to separate subtitles from the images, soundtrack and other channels of semiotic information with which they interrelate, and, therefore, the mechanical functions of subtitling from the creative processes of text production and reception. However, it is also problematic to suggest that there are not clear differences between subtitles that have been designed to demonstrate some kind of intentional interplay with other semiotic systems (usually images) to explicitly support diegetic and narrative functions, and those that have not.

It could be argued that Pedersen’s (Pedersen et al., 2022, p. 10) suggestion of the term “free form subtitles” plays into dominant ideas about subtitling being a “constrained” form of translation, as well as conventional subtitling guidelines acting as all-powerful sets of rules that restrict the creative capacities of both subtitlers and the subtitles they produce (Nornes, 1999; McClarty, 2012). The Code of Good Subtitling Practice, to name but one example, recommends that subtitles are presented in a white or yellow font (with a black drop shadow), in a plain serif typeface to ensure legibility. It further stipulates that subtitles should be limited to two lines comprising a total of 37 to 40 characters, appearing for a maximum of six seconds positioned at the bottom of the screen, where it is assumed that that less action, or less important action, will take place (Díaz Cintas & Remael 2007, p. 82). To that end, many have come to the conclusion – and criticism – that such guidelines encourage a strategy of “invisibility” (O’Sullivan, 2011; Pedersen, 2017; Romero-Fresco, 2019), whereby the aim is to make subtitles as plain and unobtrusive as possible that the viewers forget they are reading them. For others, this is something to be praised; film critic and subtitler Henri Béhar (2004, cited in O’Sullivan, 2011, p. 143) has argued, for example, that if subtitles “aren’t invisible”, this constitutes failure.

There is in fact good evidence to suggest that giving the subtitler “more freedom to create an aesthetic that matches that of the text” (McClarty 2012, p. 140) allows for subtitles to be made integral to films without disturbing their cinematic makeup. Numerous eye-tracking studies conducted on audience reception to creative subtitles (see Caffey, 2009; Brown et al., 2015; Fox, 2016; Kruger et al., 2018) conclude that the integration of the subtitles into the film’s aesthetics and visual style reduces the time and effort spent on reading, allowing them to spend more time on the images, thus creating a more effective illusion of invisibility despite displaying features that are more

visual than ever (Pollard, 2019). This, as claimed by Romero-Fresco and Fryer (2018, p. 13), consequently helps to bridge the gap between the experience of the viewers of an original work and those of its translated and/or accessible versions, while at the same time providing an exciting opportunity for collaboration and innovation between filmmakers and translators.

The aim of creative subtitles is to embed the subtitles into the audiovisual content in ways that give the impression that the filmmaker or content creator had intended for them to be there the whole time. This is reality in the case of *Night Watch*, whose director Timur Bekmambetov collaborated with Fox Searchlight, a speciality film division of 20th Century Fox and distributor of the film, on a set of subtitles for the film's English-language theatrical release that break the "conventional" rules of subtitling practice and are purposefully integrated into the film's aesthetics and visual structure. Responses from viewers, namely in reviews of the film, express a positive reaction to the subtitling experiment in general, if not the film itself; for example, Lane writes that:

The subtitles (...) are the best I have encountered. Far from palely loitering at the foot of the screen, they lurk in odd corners of the frame and, at one point, glow scarlet and then spool away, like blood in water. I trust that this will start a technical trend and that, from here on, no respectable French actress will dream of removing her clothes unless at least three lines of dialogue can be made to unwind across her midriff. (2006, online, cited in Fox, 2010, p. 86)

More specifically, here, suggestions that subtitles may either "palely loiter" or "lurk in odd corners" perhaps foreshadow the extent to which the different subtitles to be analysed in this experiment will elicit fear, if at all, which we will also be measured on a psychophysiological level. We have decided for this study to use the term "standard" to refer to those subtitles that are displayed at the bottom of the screen, usually in white letters. To acknowledge recent discussion around the term "creative subtitles", we use the term "aesthetically integrated" for those subtitles which display features that interact with the film's visuals or themes, as is explained in the next section.

3. Stimuli Selection

To examine the affective responses to both standard (STS) and aesthetically integrated subtitles (AIS), we first relied on the aesthetically integrated English-language subtitles available on the DVD version of *Night Watch*, presented with the original Russian audio. Both sets of subtitles provide an English translation but are not designed for accessibility purposes (i.e., they are not designed for D/deaf members of the audience and do not include information such as the description of sounds of the identification of speakers). According to Foerster (2010, p. 89), the "aesthetic subtitles" for the film take on two functions. One is "diegetic", in that the subtitles "serve to underline the content of the story". One subtitle design in the film that falls into this category is the use of the colour red to represent the dialogue of a vampire who frequently tries to lure the young male protagonist to her, and whose voice can only be heard by this character and the audience. The subtitle appears with varying degrees of opacity to reflect the intensity of her voice, an effect argued by McClarty (2012, p. 144) to potentially mirror the emotions of characters. In terms of special effects, these subtitles also take on the form of spurts of blood before they disappear from whichever position they

assume on the screen. In one scene, they appear in the water of a swimming pool at the same time that the young boy is having a nosebleed and seem to evaporate or splash away as the boy comes to the surface.

Whilst this particular design recurs throughout the film, others are more individualised. Some fade out slowly, instead of cueing out, or linger on screen, so as to emphasise what a character has said or foreshadow future events, in a manner that might be deemed ominous. Some subtitles flicker, blink, fade in and float across the screen in ways that accentuate the protagonist's own fragile state of mind, in particular in scenes in which he becomes mentally (and physically) unstable as various otherworldly forces overtake him and he is pulled in and out of "the gloom". In addition to showing clips with subtitles that interact with the film's diegesis, the present study also exposes participants to subtitles with a "purely stylistic purpose" (Foerster, 2010, p. 88). These include those that are positioned underneath characters or objects that are being spoken about or which come in word-by-word and resemble their movement, such as the chopping of a knife, for example. As noted by Foerster (Foerster, 2010, p. 93) in her analysis, this latter design is similar to "add-on subtitles" that are commonly used in SDH, which allow the display of a new subtitle on the screen – that is cued in sync with the speakers' utterance – before the previous one has vanished (see Díaz Cintas & Remael, 2007, p. 140). Further, the in and out times of the subtitles, i.e., the speed at which subtitles appear or leave the screen, are often determined by the composition of the image, e.g., revealed in synchrony with the opening of a fridge door and cued out as the door closes. In the context of a horror film, or a film using tactics such as jump scares to frighten the audience, for example, this type of subtitle design could be aligned with the timing of the cinematic technique and therefore contribute to its overall elicitation of fear, or other intended emotional reactions.

It is worth considering whether AIS should be selectively used in emotionally and contextually critical movie scenes, or whether they could be used throughout the duration of an entire movie (see Kabara, 2015; but see also Szarkowska & Gerber-Morón, 2018). Further research could benefit from an exploration of whether there is a threshold for habituation or audience discomfort due to AIS and further explore this effect in a dedicated publication.

For our study, twenty-two clips that display at least one of the features described above were identified with durations ranging from fifteen seconds to seventy-nine seconds. The clips were shown without any subtitles during Stage One (see below) and were shown with AIS and STS during Stage Two. In an attempt to minimize variables, rather than using other available versions of the film with STS, an STS version was recreated using the exact text of the AIS version.

4. Hypotheses

The main objective of the experiment presented in this article was to compare two subtitling delivery effects – STS versus AIS in terms of the emotional arousal of audiences, focussing specifically on fear. Taking the results of previous experimental studies conducted on the reception of films with AIS in terms of immersion, cognitive load, aesthetic experience, etc., we hypothesized that AIS would result

in higher ratings for emotional and viewing experience than STS. Along these results, we also hypothesized that we would record higher positively rated arousal across the psychophysiological measures used (SCR and HR) for AIS than for STS. Whilst previous applications of this methodology in AVT reception studies have produced data that is somewhat inconclusive, we hypothesized that focussing on the elicitation of one single emotion would show clearer results.

5. Experiment

5.1. Stage One: Stimuli Preselection

Aims: The aim of this stage was to assess participants' responses to select the most characteristic clips for a fantasy-thriller audiovisual genre/content from the film *Night Watch*.

Participants: A power calculation based on a within-subject design and small effect sizes ($d = .2$; Faul et al., 2009) given the specific experimental trial contour ($k = 22$; see Baker et al., 2020) was performed. The result revealed that forty-eight participants would be required for $P(1-\beta) \geq .9$; ($p \leq .05$; $P(H_0) \geq .9$; $B < .33$; $d [\geq -0.1, \leq .1]$) (see also Kruschke & Liddell, 2018). Fifty-one volunteers participated in this stage. The exclusion criteria were that participants should not have seen the film *Night Watch* or seen short scenes, trailer advertisements and movie-related commentaries of the film in any format. The inclusion criteria were that participants should be native monolingual (see Bidelman & Heath, 2019) non-dyslexic English speakers who were not taught or self-educated in Russian or a related language of Slavic origin (see Grenoble, 2010).

Participants were enquired and assessed for clinical and subclinical conditions that could bias their performance during the experiment (see Sultana et al., 2021). They should not have a current or previous DSM-5 Axis I (clinical disorders) or II (personality disorders) diagnosis (American Psychiatric Association, 2013) or a diagnosis for dyslexia; self-report (Y/N). To further explore these self-reports, the participants were screened with the Somatic and Psychological Health Report Questionnaire (SPHRQ; Hickie et al., 2001; Berryman et al., 2012), with the Stressful Life Events Screening Questionnaire (SLESQ; Goodman et al., 1998; Gray et al., 2004; Allen et al., 2015) and the Behavior Rating Inventory of Executive Functions Questionnaire in Dyslexia Questionnaire (BRIEF-P/T; Akyurek et al., 2018). To assess whether participants could respond to the emotional context of the presented clips, they were assessed with an online Alexithymia Questionnaire (AQ, 2020; see Ridout et al., 2020). Three participants were excluded from the analyses due to having scores indicating possible Alexithymic traits (see Thompson, 2007). No participant was excluded based on their SPHRQ, SLESQ and BRIEF-P/T assessments.

The participants were further assessed with emotional regulation and sensitivity to fear questionnaires that could reveal personality traits that could bias their performance during the experiment (see Clasen et al., 2020). They were assessed with the Cognitive Emotional regulation Questionnaire (CERQ-short; Garnefski & Kraaij, 2006) and the Trait Fear Scale (TFS; Kramer et al.,

2020). Data from participants who indicated outlier values ($\pm 1.5 \times \text{InterQuartile Range}$) would be considered subject to exclusion. No participant was excluded based on these assessments (see Table 1; see Tsikandilakis et al., 2021a). Finally, participants were invited through the University of Nottingham participation scheme and assessed for achieving a proportional sampling of viewer characteristics (see Chen et al., 2001). They were asked to report how many films and TV shows they watch on average per week. To achieve a proportional sampling, the participants were included in three groups with equal n and gender distributions. One group ($n = 16$) reported seeing one to four hours of movies or series per week, one group five to eight hours and one group nine to twelve hours per week. Conversely participants were provided with a list of commercially available movies, series and animations with aesthetically integrated subtitles and asked to select the ones they had previously seen. An open-ended question item also allowed participants to state if they had seen a video with AIS that was not part of the list. For participant habituation and exposure experience to subtitled content, see Appendix 1. To achieve a proportional sampling again three groups with equal gender distributions were formed. One group included participants who had watched none to two videos with AIS, one group included participants who had watched three to five and one group included participants who had watched six to eight videos with AIS. The combination of these nine groups provided an overall population sample that consisted of forty-eight participants (twenty-four female; see also Table 1). The experiment in Stage One was approved by the Ethics Committee of the School of Cultures, Languages and Area Studies of the University of Nottingham.

Table 1

Participant Characteristics Overall and per Gender for Stage One

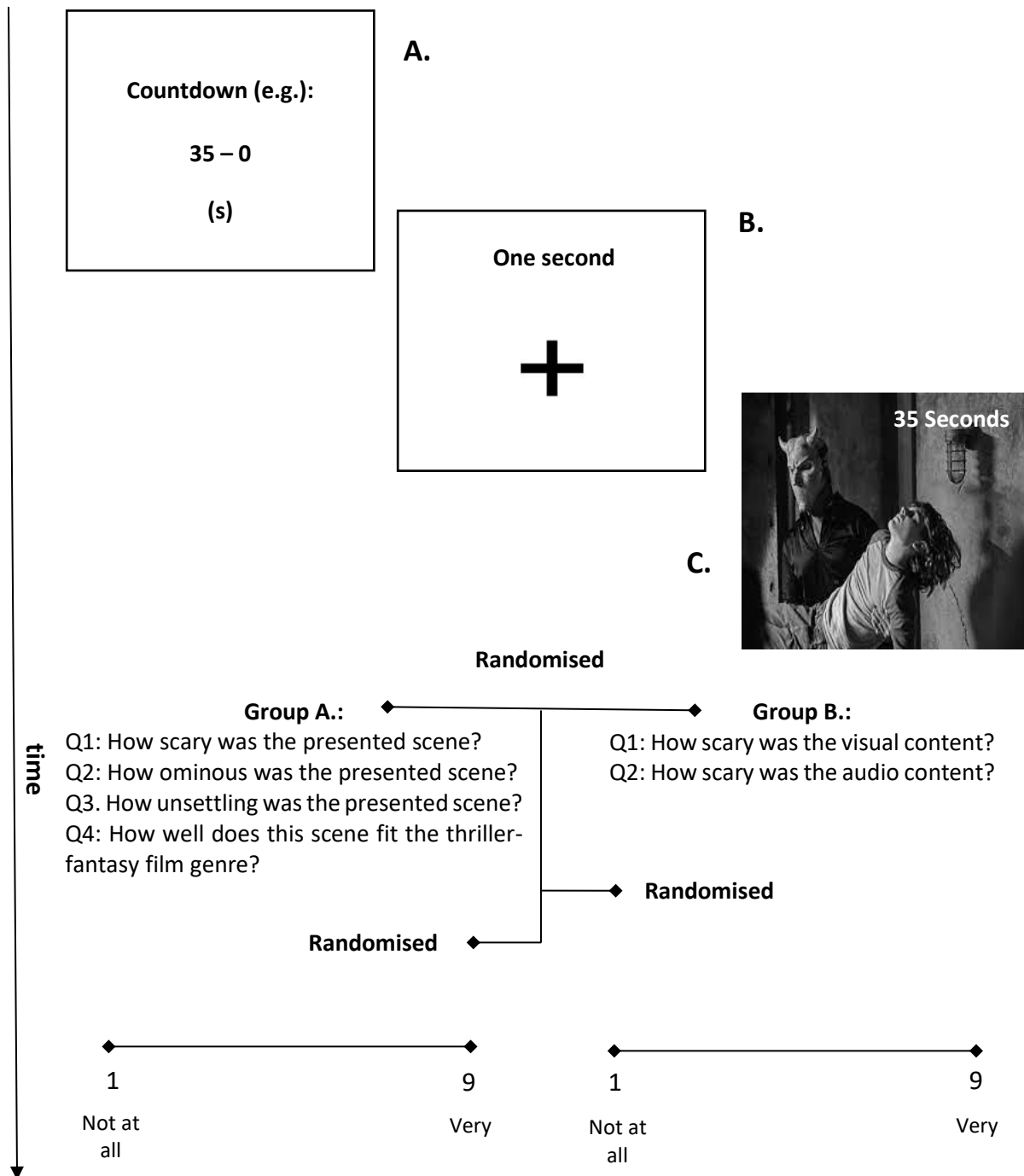
Participants	<i>n</i>	Age	SPHRQ	AQ	BRIEF-P/T	SLESQ	CERQ-short		FTQ
		Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	CR	ES
Overall	48	24.12 (2.42)	1.11 (.07)	61.03 (2.09)	45.67 (9.12)	1.23 (.62)	26.71 (3.04)	8.33 (1.16)	1.78 (.21)
Male	24	24.43 (2.17)	1.09 (.06)	62.07 (2.35)	46.18 (9.09)	1.24 (.69)	26.79 (3.11)	8.31 (1.34)	1.77 (.25)
Female	24	23.81 (2.67)	1.13 (.08)	59.96 (1.75)	44.72 (9.14)	1.22 (.61)	26.62 (2.92)	8.35 (1.01)	1.79 (.27)
T-test and Bayesian Analyses for Gender Comparisons									
P-value						.79	.75	.81	.72
(Cohen's <i>d</i>)						(.03)	(.05)	(-.02)	(-.07)
Bayes Factor						.09	.08	.19	.04

Note. This table includes participant *n* and age per gender. It also includes mean and standard deviation percentiles for the cognitive emotional regulation questionnaire (CERQ-short) with scores for cognitive re-appraisal (CR) and emotional suppression (ES) and for the trait fear questionnaire, the AQ, the BRIEF-P/T and the SLESQ mean average of forty-four, sixty and thirteen items respectively. In the bottom of the table frequentist and Bayesian comparisons between emotional questionnaire reports for male and female participants (see Tsikandilakis et al., 2021b). We did not present relevant comparison per group categories because proportional sampling is designed to provide sample diversity (Chen et al., 2001). A complete account of information and characteristics per group and group of combinations of participants can be found in Appendix 1.

Stimuli: The audiovisual stimuli comprised the twenty-two clips from the fantasy-thriller film *Night Watch*. The stimuli were presented in a HD (1080p) monitor set at 60Hz (Refresh Rate = 16.67 ms) with the colour contrast, brightness and luminosity values that were in-built in the commercially available film copies. The clips were presented from the beginning to the end of each clip. Their range was from fifteen to seventy-nine seconds with a mean duration of 39.13 seconds (*SD* = 17.97). The clips were presented without any form of subtitling to evaluate the baseline audiovisual ratings for each scene before assessing the impact of subtitling methods during Stage Two (see Cacioppo et al., 2007, pp. 164–173).

Procedure: We used a within-subjects design to select the most characteristic clips for a fantasy-thriller audiovisual genre/content (see Van Hoecke et al., 2022; see also particularly the defining characteristics of thriller-fantasy films as defined by Oliver & Sanders, 2004; pp. 118–124). We matched the experimental design parameters for Stage One with the experimental design parameters in Stage Two. Every presentation started with a numeric countdown at fixation using Calibri (body) bold size sixteen font. The countdown was in seconds and was equal in duration with the subsequently presented clip (see Stage Two). After the countdown a fixation cross was presented for one second and then an audiovisual clip was presented without any subtitles with order of clip presentation randomised. After the presented clip, a five-second blank screen interval was presented and then the participants were presented with two blocks of questions that included a set of question-assessments presented separately and one at a time. Participants were asked to respond to these questions using the mouse. In Block A the participants were asked to rate from one (not at all) to nine (very) how scary the presented clip was, how ominous the presented clip was, how unsettling the presented clip was and how well the clip fits the thriller-fantasy film genre. In Block B participants were asked to rate from one (not at all) to nine (very) how scary the visual content was and how scary the audio content was. Each block was presented separately and randomly and the questions within each block were presented randomly in each trial (see Figure 1).

Figure 1
Experimental Sequence for Stage One



Note. Example experimental sequence for Stage One including A. the pre-stimulus countdown set at an example thirty-five seconds, B. the one-second fixation cross, and C. an open-source black-and-white illustration from a fantasy-thriller film (all included clips were presented in colour). At the bottom of the figure are presented the two groups of questions including the question-assessments and an example Likert scale as included in the experimental design. The order of the question groups and the order of the question-assessments in each group were randomised in each trial.

Psychometrics: Typically, null hypothesis significance testing (NHST) is used for stimuli assessment in audiovisual studies (see Gross & Levenson, 1995). In NHST, the researcher examines whether they can reject or fail to reject the null hypothesis; that, for example, reporting substantial differences between two or more variables is true. If the probability for erroneously failing to reject the null hypothesis is equal or lower to an arbitrary threshold (traditionally less than 5%; see Lakens et al., 2018), the researcher can reject the null hypothesis and the result is termed significant (Banerjee et al., 2009). If the probability is more than 5%, the researcher fails to reject the null hypothesis and the result is termed nonsignificant. Within the context of preselecting audiovisual stimuli this has resulted in comparing different stimuli to each other and selecting the ones that provide significant effects for having the highest ratings for a given genre/context. In the current study we wanted to select the most characteristic and not the most intense stimuli for a thriller-fantasy genre/context (see Brosch et al., 2010). To address this objective, we used Bayesian statistics. Bayesian analysis can provide direct evidence for ratings being within an a-priori defined range. Bayesian analysis requires a lower and an upper bound, the standard error of the population sample and a simple deduction of the sample mean from the theoretical comparison mean to provide a Bayes Factor (BF). Using these parameters we can calculate a BF that will signify at $BF < .33$ direct evidence for the null, in the current design, that the ratings are within a-priori defined range, at $.33 < BF < 3$ that the results are inconclusive and at $BF > 3$ evidence for the likelihood of the data being observed under the alternate hypothesis; in the current design, that the reported ratings are outside our desired range (Dienes, 2014, 2015, 2016).

Results and Discussion: In the current design, given the lack of research precedence in the current area, we employed a uniform distribution for Bayesian analysis (see Liu & Aitkin, 2008), we treated our rating data – as suggested by recent topical reviews (see Wagermakers et al., 2018a, 2018b; Bürkner & Vuorre, 2019) – as sequential ordinal data and implemented a motivated by minimum effect sizes of interest characteristics Bayesian analysis ($f = .05$; $d = .1$; see Dienes, 2019, 2021). We defined the a-priori characteristic value for ominous, scary (Group A. questions), unsettling, and video-scary and audio-scary at 7 with lower bounds and higher bounds respectively at -1 (6) and +1(8) (Dienes, 2019, pp. 366–369). For the question “How well does this scene fit the thriller-fantasy film genre?” we defined a desirable value at 9 with a lower bound set -2 and a higher bound set at 0. We were able to find seven (7/22 overall) clips that provided Bayesian evidence for meeting these criteria. The clips had a range duration from 15 to 79 seconds with $M = 35.86$ ($SD = 21.36$). The ratings and Bayesian analyses for the selected clips – with clips named in chronological order as presented in *Night Watch* – are illustrated in Table 2.

Table 2

Ratings and Bayesian Inference Analyses for the Selected Clips

Mean (SD)						
	Ominous	Scary (A.)	Unsettling	Fitting	Video-Scary (B.)	Audio-Scary (B.)
CLIP 1	7.03 (.21)	7.04 (.22)	7.01 (.19)	8.96 (.23)	7.03 (.19)	7.01 (.18)
CLIP 2	7.02 (.22)	7.04 (.21)	7.04 (.22)	8.96 (.29)	7.04 (.18)	7.03 (.21)
CLIP 3	7.03 (.18)	7.03 (.19)	7.03 (.21)	8.95 (.28)	7.03 (.20)	7.04 (.19)
CLIP 10	6.98 (.21)	6.97 (.22)	6.96 (.20)	8.94 (.21)	6.96 (.24)	6.98 (.22)
CLIP 12	7.01 (.19)	7.04 (.21)	7.04 (.24)	8.97 (.19)	7.05 (.21)	7.03 (.21)
CLIP 14	7.03 (.18)	7.04 (.19)	7.04 (.21)	8.97 (.22)	6.99 (.25)	7.04 (.28)
CLIP 17	7.03 (.22)	7.04 (.17)	7.02 (.25)	8.96 (.19)	6.98 (.25)	7.02 (.25)
Bayes Factor (SE)						
	Ominous	Scary (A.)	Unsettling	Fitting	Video-Scary (B.)	Audio-Scary (B.)
CLIP 1	.08 (.03)	.19 (.03)	.03 (.01)	.18 (.03)	.11 (.02)	.05 (.02)
CLIP 2	.04 (.03)	.21 (.03)	.19 (.03)	.08 (.03)	.15 (.02)	.08 (.03)
CLIP 3	.11 (.02)	.14 (.02)	.08 (.03)	.14 (.03)	.21 (.03)	.19 (.02)
CLIP 10	.27 (.03)	.08 (.03)	.21 (.03)	.27 (.03)	.21 (.03)	.19 (.03)
CLIP 12	.03 (.02)	.21 (.03)	.17 (.03)	.07 (.02)	.57 (.03)	.11 (.03)
CLIP 14	.09 (.02)	.19 (.02)	.23 (.03)	.05 (.03)	.03 (.03)	.27 (.03)
CLIP 17	.08 (.03)	.14 (.02)	.19 (.03)	.09 (.02)	.05 (.03)	.19 (.03)

Note. Ratings for each question item and Bayes Factors (BF) with standard deviation (SE) for the selected clips from Stage One. The current clips were the only clips out of the twenty-two overall from *Night Watch* that met the criteria for further analyses in Stage Two. Clip 12 for item “Video-Scary” provided a Bayesian trend ($BF = .57$) for meeting the requirements of the analyses. Scary A. refers to the first group of questions, and Video-Scary and Audio-Scary B. refer to the second group of questions (see Figure 1).

5.2. Stage Two: Aesthetically Integrated vs Standard Subtitles

Aims: The aim of this stage was to assess participants’ physiological and rating responses to the clips we selected during Stage One from *Night Watch* when these were presented with AIS and standard subtitles.

Participants: No data in this stage provided evidence when using Kolmogorov-Smirnov and Shapiro-Wilks tests for violations of normality (see Thode, 2002). A power calculation based on a between-subjects design and small effect sizes ($d = .2$) and between-subject experimental trial contour ($k = 7$) was performed. The result revealed that 180 participants would be required for $P(1-\beta) \geq .9$; ($p \leq .05$; $P(H_0) \geq .9$; $B < .33$; $d [\geq -0.1, \leq .1]$). One-hundred-and-eighty-seven participants volunteered to take part in the current stage. The inclusion and exclusion criteria, and clinical and subclinical and participant assessment questionnaires required for Stage Two were identical with Stage One. To achieve a proportional sampling, the same groups as Stage One were comprised for each group of

Stage Two. One group ($n = 30$) reported seeing one to four hours of films or TV shows per week, one group five to eight hours and one group nine to twelve hours per week. Again, participants were provided with a list of commercially available films, series and animations with aesthetically integrated subtitles and asked to select the ones they had previously seen. Again, an open-ended question item also allowed participants to state if they had seen a video with AIS that was not part of the list. We formed three groups. One group included participants who had watched none to two videos with AIS, one group included participants who had watched three to five and one group included participants who had watched six to eight videos with AIS (see Appendix 2). Three participants were excluded from the analyses due to scores that indicated Alexithymic traits. Three participants were excluded from the analyses as outliers for having seen more than thirteen ($M = 15$; $SD = 1$) films with AIS. One participant was excluded due to movement artefacts during the psychophysiological assessment (see Cacioppo et al., 2007, pp. 193–205). The final population sample consisted of 180 participants (ninety female). From this final sample, with equal proportional sampling distributions, one group included ninety participants (forty-five female; AIS Group) and one group included ninety participants (forty-five female; Standard Subtitles (STS) Group; see Table 3). No participants from Stage One were included in Stage Two. The assignment of the participant per group was made using pseudorandomised coding (Python) based on acronyms (e.g., F_G1_G2) that involved gender and group affiliation. The experiment in Stage Two was approved by the Ethics Committee of the School of Cultures, Languages and Area Studies of the University of Nottingham.

Table 3

Participant Characteristics in Stage Two Overall and per Group

Participants	Overall (n)	Female (n)	Age	SPHRQ	AQ	BRIEF-P/T	SLESQ	CERQ-short		FTQ
			Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)
								CR	ES	TFS
Overall	180	90	25.28 (2.04)	1.12 (.09)	60.15 (2.09)	45.17 (8.11)	1.24 (.51)	25.6 (3.39)	8.11 (1.06)	1.59 (.23)
Group One (AIS)	90	45	25.13 (2.12)	1.07 (.09)	60.04 (2.01)	45.18 (8.09)	1.25 (.56)	25.57 (3.31)	8.1 (1.01)	1.62 (.21)
Group Two (STS)	90	45	25.42 (1.92)	1.18 (.09)	60.26 (2.17)	45.16 (8.13)	1.23 (.46)	25.63 (3.47)	8.12 (1.11)	1.56 (.25)
T-test and Bayesian Analyses Gender Comparisons										
P-value (Cohen's d)							.92 (.01)	.82 (- .03)	.79 (- .02)	.71 (- .08)
Bayes Factor							.03	.09	.21	.27

Note. This table includes participant n and age per group. It also includes mean and standard deviation percentiles for the cognitive emotional regulation questionnaire (CERQ-short) with scores for cognitive re-

appraisal (CR) and emotional suppression (ES) and for the trait fear questionnaire, the AQ, the BRIEF-P/T and the SLESQ mean average of forty-four, sixty and thirteen items respectively. In the bottom of the table the reader can find frequentist and Bayesian statistical outcomes between the two groups. Information and characteristics per gender, group and group of combinations of participants can be found in Appendix 2.

Stimuli: The audiovisual stimuli comprised the seven clips from the fantasy-thriller film *Night Watch* selected during Stage One (see Stage One: Results and Discussion). One group of stimuli allocated to one group of participants included standard subtitles and one group of stimuli allocated to the other group of participants included AIS.

Procedure: The procedure was similar to Stage One with a few variations. These variations included that one group was allocated AIS stimuli only (Group One) and one group was allocated STS stimuli only (Group Two). Both groups were assessed for psychophysiology during the experiment (see Stage Two: Psychophysiology). In both groups, after the presentation of the AIS or STS clips, participants were asked to rate from one (not at all) to nine (very) from two categories of question sets. One set involved how scary the presented clip was, how ominous the presented clip was and how unsettling the presented clip was. One set involved how the subtitles affect their viewing experience, how immersed they were during the scene presentation, how well they understood the context of the presented clip and how they would rate their viewing experience. The order of the question sets and the questions within each set were presented in randomised order in each trial (see also Figure 1).

Psychophysiology: Skin conductance (SCR) and heart rate (HR) were used to assess physiological responses. SCR were measured from the left hand (index/first and middle/second fingers) of each participant using disposable Ag/AgCl gelled electrodes. The signals were received by a BIOPAC System, EDA100C in units of microsiemens (μS) and recorded in AcqKnowledge (Braithwaite et al., 2013). HR was measured via a single finger sensor from the left hand (ring/third finger). The signal was received by a BIOPAC System, PPG100C in units of beats per minute (bpm) and recorded in AcqKnowledge.

We took two different measurements for each presented experimental sequence. First, we calculated the tonic SCR and HR baseline during each scene (see Figure 1: C.), and then compared it between AIS and STS clips to acquire an overall sense of physiological changes during these two conditions (see Cacioppo et al., 2007, pp. 164). Second, we used parallel input for phasic event-related AIS and STS instances that marked their onset, duration, offset and markers for SCR up to three seconds (see Cacioppo et al., 2007, pp. 164–167) and for HR up to five seconds post-scene-offset (see Cacioppo et al., 2007, pp. 187–191) in each scene. We derived for both assessments phasic event-related responses very conservatively as an unambiguous increase ($\text{SCR} \geq \pm .01 \mu\text{S}$; $\text{HR} \geq \pm 1 \text{ bpm}$; the latter derived from a combination of at-least three subsequent event-related sequential beats) (see Cacioppo et al., 2007, p. 188) with respect to the average tonic baseline of each AIS and STS scene (see Cacioppo et al., 2007, pp. 168–173). The raw signals for both measures were processed using the Derive Phasic from Tonic and manual Dirac's delta (δ) functions (see Balakrishnan, 2003). The data did not require additional smoothing, filtering or transformations (Braithwaite et al., 2013, p. 10–12). Non-responders for physiological changes were included in the

data analysis (see Van Der Ploeg et al., 2017; pp. 143–147). The methods used for psychophysiological assessment were intended also as an assessment index for further topical research and, therefore, they were defined very conservatively and with as unbiased comparison baseline periods as feasible (see Cacioppo et al., 2007; pp. 164–167, 187–191; see also Tsikandilakis et al., 2018; Tsikandilakis et al., 2019, 2020a, 2020b).

Results and Discussion; Participant Responses: To compare the participant responses to AIS and STS clips, we ran a series of Bonferroni corrected ($p \leq .01$) between-subjects t-tests. We also included Bayesian statistics in the analyses with full details for replication based on a uniform distribution and information about the mean difference (MD), the standard error (SE), the lower (LB) and higher bounds (HB) and Bayes Factors (BF) for each statistical report, as reported in the stimuli selection stage (see Wagemakers et al., 2018a, 2018b; Bürkner & Vuorre, 2019; Dienes, 2019, 2021).

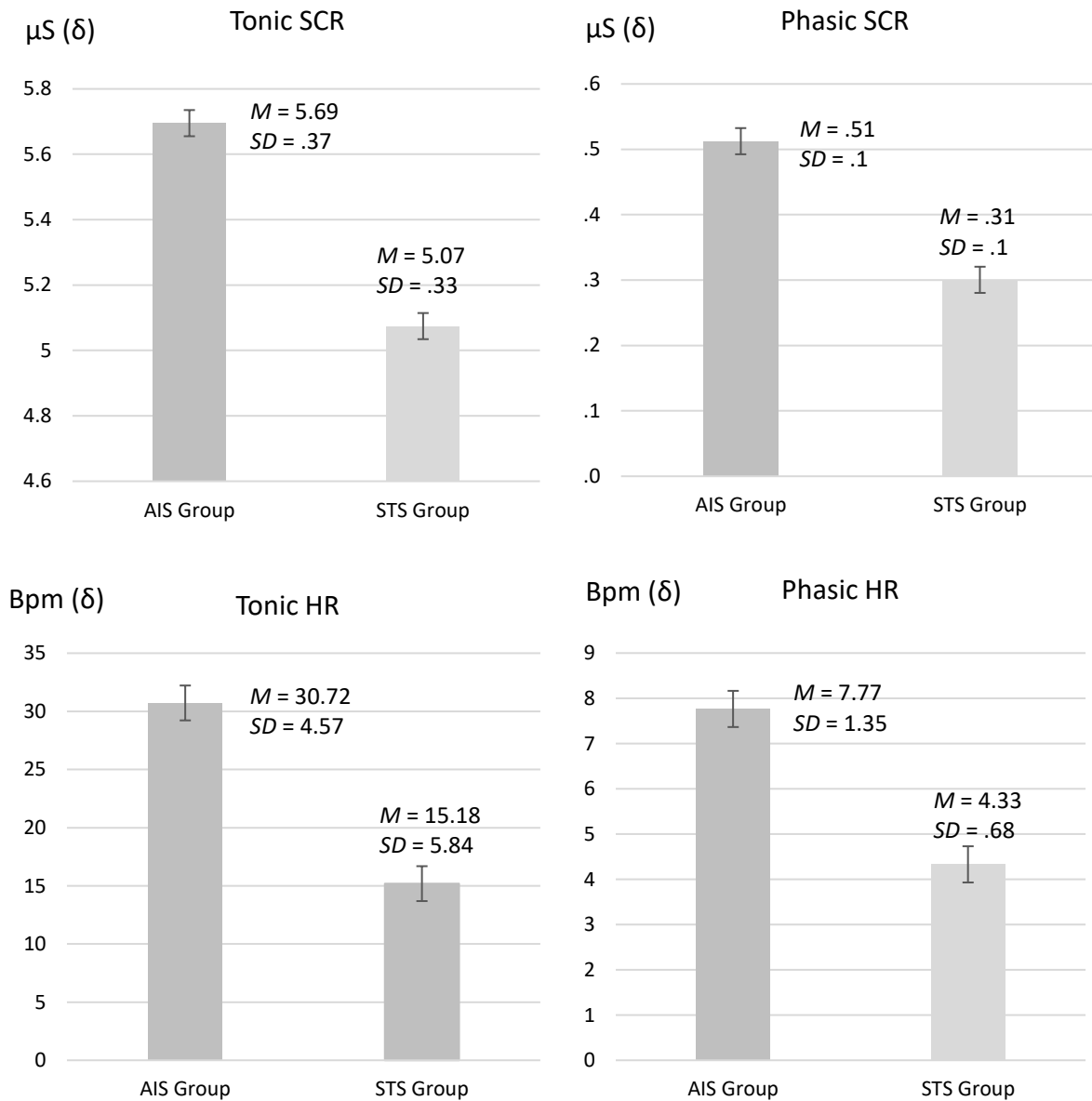
Participants in the AIS group ($M = 7.99$, $SD = .32$) rated the presented clips as being scarier than participants in the STS group ($M = 6.26$, $SD = .18$; $t(88) = 42.57$; $p \leq .001$; $d = 6.66$; $MD = 1.73$, $SE = .02$, $LB = -1$ & $HB = +1$; $BF = +\infty$). Participants in the AIS group ($M = 7.69$, $SD = .26$) rated the clips higher for how ominous they were compared to the STS group ($M = 6.34$, $SD = .18$; $t(88) = 31.07$; $p \leq .001$; $d = 6.04$; $MD = 1.35$, $SE = .02$, $LB = -1$ & $HB = +1$; $BF = +\infty$). Participants in the AIS group also rated the clips as more unsettling ($M = 7.55$, $SD = .3$) compared to the STS group ($M = 6.18$, $SD = .21$; $t(88) = 37.3$; $p \leq .001$; $d = 5.29$; $MD = 1.37$, $SE = .03$, $LB = -1$ & $HB = +1$; $BF = +\infty$). Conversely, AIS group participants reported that the subtitles increased the quality of their viewing experience ($M = 8.01$, $SD = .35$) compared to the STS group ($M = 5.06$, $SD = .33$; $t(88) = 55.67$; $p \leq .001$; $d = 8.67$; $MD = 3.04$, $SE = .03$, $LB = -1$ & $HB = +1$; $BF = +\infty$). The same effect was reported for AIS participants for immersion in the clip context ($M = 7.75$, $SD = .27$) compared to the STS participants ($M = 6.12$, $SD = .42$; $t(88) = 31.26$; $p \leq .001$; $d = 4.62$; $MD = 2.95$, $SE = .03$, $LB = -1$ & $HB = +1$; $BF = +\infty$). Finally, AIS group participants rated their overall viewing experience higher ($M = 8.15$, $SD = .23$) than STS group participants ($M = 6.39$, $SD = .33$; $t(88) = 41.76$; $p \leq .001$; $d = 6.19$; $MD = 1.76$, $SE = .03$, $LB = -1$ & $HB = +1$; $BF = +\infty$). These results suggest that AIS subtitles were higher for all viewing quality assessments compared to STS subtitles.

Results and Discussion; Psychophysiology: Tonic (overall) arousal for SCR was higher for the AIS group compared to the STS group ($t(88) = 27.89$; $p \leq .001$; $d = 7.78$; $MD = .62$, $SE = .55$, $LB = -1$ & $HB = +1$; $BF = +\infty$). Phasic (post-stimulus offset) SCR arousal was also higher for the AIS group compared to the STS group ($t(88) = 15.79$; $p \leq .001$; $d = 2.4$; $MD = .21$, $SE = .03$, $LB = -.01$ & $HB = +.01$; $BF = +\infty$). For HR, the AIS group also responded with higher tonic (overall) arousal to the presented clips compared to the STS group ($t(88) = 20.95$; $p \leq .001$; $d = 7.12$; $MD = 15.53$, $SE = .53$, $LB = -10$ & $HB = +10$; $BF = +\infty$). The AIS group also experienced higher phasic (post-stimulus offset) HR arousal compared to the STS group ($t(88) = 23.25$; $p \leq .001$; $d = 5.72$; $MD = 3.43$, $SE = .11$, $LB = -2.5$ & $HB = +2.5$; $BF = +\infty$).

see Figure 2). These findings suggest that AIS group participants experienced higher, positively rated during the self-report tasks, physiological arousal compared to STS group participants.²

Figure 2

Tonic and Phasic SCR Responses for AIS and STS Groups in Stage Two



Note. Psychophysiological, SCR (µS) and HR (bpm), Dirac's delta (δ; Balakrishnan, 2003) responses by AIS and STS participant groups in responses to pre-selected clips. Mean and standard deviation are included in the graph label. Bars show ±2 standard errors of the mean.

² The materials for the current study have been open access at <https://osf.io/q9hu7>.

6. Discussion and Conclusion

In this manuscript, we presented two studies. In the first study, we pre-selected the most characteristic clips from the film *Night Watch* that fit a fantasy-thriller genre/context. In the second study, we showed that participants reported that these pre-selected clips were higher for watching experience and fantasy-thriller movie characteristics, such as how scary, unsettling and ominous they were, and experienced higher positively rated physiological responses as measured by SCR and HR when they involved AIS subtitles compared to when they involved STS subtitles.

Our first hypothesis that AIS would result in a more intense emotional arousal as reported by participants than STS was confirmed. Our second hypothesis that we would see higher values across the psychophysiological measures used (SCR and HR) for AIS than STS was also confirmed. It should be noted again here that our study is limited to one single emotion, in the context of clips taken from one single film. Naturally, further studies ideally need to allow for an assessment of viewer endurance and overall experience during prolonged exposure to AIS. Further research could benefit from an exploration of other emotions, different types of material, and the factoring in of viewers' different ages and levels of habituation with regards to subtitles. *Night Watch* provided an excellent testing ground because the AIS were produced with input from the director himself, whose authorial intent naturally provides a certain level of authority. Another limitation that comes as a consequence though is that arousal is not narrowed to a single particular feature of the subtitles (position, colour, visual effect). Rather, our study considers what two different sets of subtitles achieves.

It should also be taken under consideration that the current study recruited a significant population sample that conferred high statistical power. The participants all volunteered to take part in the current study. This could suggest that the current study was of personal interest to the participants, or that the subject of the study could be a motivational factor for participating and being observed as contributing volunteers to an area of personal interest. This effect, more widely known as the Hawthorne effect, is suggested to be an inevitable characteristic of studies involving psychophysiological assessments, that require, for ethical reasons, explicit pre-briefing to the participants. Despite these, this effect merits mention in the current study because it could have influenced the responses of the current population sample. Participants could have assumed expectancies to positively rate the AIS as a more contributing subtitling method. It must also be stated that the automatic and involuntary psychophysiological responses reported in this research, are considered to occur irrespective of conscious inhibition and self-presentation biases and are not considered as outcomes that could be influenced by the Hawthorne effect (see McCambridge et al., 2014).

One unexpected finding of the study was that the participants reported more intense emotional arousal when watching the clips without subtitles compared to clips with STS. Whilst this is not entirely counter-intuitive, it provides some evidence that the presence of STS (again, with the limitations noted above – in the context of short clips from *Night Watch*, in the context of fear) is less conducive to the desired viewing experience than having no subtitles at all.

Our study therefore feeds into the growing body of research showing that AIS have the potential to improve the overall film viewing experience, though, to do so successfully, as Romero-Fresco (2019) points out, greater collaboration between the film and AVT industries and investment in their resources is needed.

The absence of bespoke tools, in particular, has been identified as an important area of development for aesthetically integrated subtitling to become easier to implement (Mével, 2020; Mével et al., 2022) and, therefore, to become a more widespread practice. Currently, no specialist software exists for producing AIS as efficiently and spontaneously as STS software allows, and STS software, for its part, typically does not have a range of features wide nor sophisticated enough to produce well-designed, professional looking AIS. As a result, a great deal of time and effort is required to design, produce, place and edit the specific subtitle and its effects, as well as a high degree of competence with video compositing and animation software if users want to create more complex designs. Conducting experimental studies with audiences more broadly represent an important step in the development of AIS as a practice, both in terms of testing and evaluating different techniques, effects and styles, and producing taxonomies and descriptive models that may streamline the creative process for practitioners, as well as for demonstrating the advantages and opportunities that AIS offers to viewers, translators and most crucially filmmakers, whose investment is needed to mainstream the practice and widen access to it altogether.³ To this end, having delivered the first to our knowledge empirical evidence that both participant ratings and psychophysiological responses for the quality of viewing experience are higher for AIS compared STS in a fantasy/thriller context, using state-of-the-art methodological and statistical assessments, we are called to explore these findings further, for additional emotions, for diverse audiences, and for diverse film genres. We are called to provide easy-to-use software for their implementation. We have shown the “what” and are now called to explore the “where”, “when” and “how to” in order to further our understanding of this potential paradigm-shift that could shape the future of audiovisual translation.

³ On this last point, see also Romero-Fresco (2019). The effects of AIS (that may be designed simply to provide translation as is the case in *Night Watch*, or on the contrary may be designed specifically with accessibility in mind) on D/deaf members of the audience also remain to be investigated.

References

- Akyurek, G., Bumin, G., & Karabulut, E. (2018). Reliability and validity of the behavior rating inventory of executive functions questionnaire in dyslexia. *Archives of Physical Medicine and Rehabilitation, 99*(10), e20.
- Alexithymia (2022, May 10). *Alexithymia: Emotional blindness*.
<https://www.alexithymia.us/alexithymia-questionnaire-online-test>
- Allen, J. G., Madan, A., & Fowler, J. C. (2015). Reliability and validity of the Stressful Life Events Screening Questionnaire among inpatients with severe neuropsychiatric illness. *Bulletin of the Menninger Clinic, 79*(3), 187–202.
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Author.
- American Psychological Association. (2019). *Publication Manual of the American Psychological Association* (7th ed.). Author.
- Baker, D. H., Vilidaite, G., Lygo, F. A., Smith, A. K., Flack, T. R., Gouws, A. D., & Andrews, T. J. (2020). Power contours: Optimising sample size and precision in experimental psychology and human neuroscience. *Psychological methods*.
- Balakrishnan, V. (2003). All about the Dirac delta function (?). *Resonance, 8*(8), 48–58.
- Banerjee, A., Chitnis, U. B., Jadhav, S. L., Bhawalkar, J. S., & Chaudhury, S. (2009). Hypothesis testing, type I and type II errors. *Industrial Psychiatry Journal, 18*(2), 127–131.
- Bassnett, S., Venuti, L., Pedersen, J., & Hostová, I. (2022). Translation and creativity in the 21st century. *World Literature Studies, 14*(1), 3–17. <https://doi.org/10.31577/WLS.2022.14.1.1>
- Bauer, E. P. (2022). Sex differences in fear responses: Neural circuits. *Neuropharmacology, 109*298.
- Bekmambetov, T. (Director). (2004). *Night Watch* [Film]. Channel One Russia.
- Bergstrom, J. R., & Schall, A. (Eds.). (2014). *Eye tracking in user experience design*. Elsevier.
- Berryman, C., McAuley, J. H., & Moseley, L. G. (2012). Sphere 12 screening questionnaire. *Journal of Physiotherapy, 58*(4), 273.
- Bidelman, G. M., & Heath, S. T. (2019). Enhanced temporal binding of audiovisual information in the bilingual brain. *Bilingualism: Language and Cognition, 22*(4), 752–762.
- Bisson, M. J., Van Heuven, W. J. B., Conklin, K., & Tunney, R. J. (2014). Processing of native and foreign language subtitles in films: An eye tracking study. *Applied Psycholinguistics, 35*(2), 399–418. <https://doi.org/10.1017/S0142716412000434>
- Black, S. (2020). Could integrated subtitles benefit young viewers? Children’s reception of standard and integrated subtitles: A mixed methods approach using eye tracking. *Perspectives, 30*(3), 503–519. <https://doi.org/10.1080/0907676X.2020.1849324>
- Boone, W. J., & Staver, J. R. (2020). *Advances in Rasch analyses in the human sciences*. Springer.
- Bos, M., Jentgens, P., Beckers, T., & Kindt, M. (2013). Psychophysiological response patterns to affective film stimuli. *PloS One, 8*(4), 1–8. <https://doi.org/10.1371/journal.pone.0062661>
- Braithwaite, J. J., Watson, D. G., Jones, R., & Rowe, M. (2013). A guide for analysing electrodermal activity (EDA) & skin conductance responses (SCRs) for psychological experiments. *Psychophysiology, 49*(1), 1017–1034.
- Brooks, M., & Armstrong, M. (2014). *TVX2014 short paper – enhancing subtitles*. BBC.
<http://www.bbc.co.uk/rd/blog/2014-10-tvx2014-short-paper-enhancing-subtitles>
- Brosch, T., Pourtois, G., & Sander, D. (2010). *The perception and categorisation of emotional stimuli: A review* (pp. 76–108). Psychology Press.

- Brown, A., Jones, R., Crabb, M., Sandford, J., Brooks, M., Armstrong, M., & Jay, C. (2015). *Whitepaper WHP305. Dynamic subtitles: The user experience*. BBC. <http://downloads.bbc.co.uk/rd/pubs/whp/whp-pdf-files/WHP305.pdf>
- Brumbaugh, C., Kothuri, R., Marci, C., Siefert, C., & Pfaff, D. (2013). Physiological correlates of the big 5: autonomic responses to video presentations. *Applied Psychophysiology and Biofeedback*, 38(4), 293–301. <https://link.springer.com/article/10.1007/s10484-013-9234-5>
- Bürkner, P. C., & Vuorre, M. (2019). Ordinal regression models in psychology: A tutorial. *Advances in Methods and Practices in Psychological Science*, 2(1), 77–101.
- Cacioppo, J. T., Tassinary, L. G., & Berntson, G. (Eds.). (2007). *Handbook of psychophysiology*. Cambridge University Press.
- Caffrey, C. (2009). *Relevant abuse? Investigating the effects of an abusive subtitling procedure on the perception of TV anime using eye tracker and questionnaire* [Doctoral dissertation, Dublin City University]. Doras (DCU Online Research Access Service) <http://doras.dcu.ie/14835/>
- Carroll, N. (2008). *The philosophy of motion pictures*. Blackwell.
- Chaume, F. (2004). Film studies and translation studies: Two disciplines at stake in audiovisual translation. *Meta*, 49(1), 12–24. <https://doi.org/10.7202/009016ar>
- Chen, T. Y., Tse, T. H., & Yu, Y. T. (2001). Proportional sampling strategy: A compendium and some insights. *Journal of Systems and Software*, 58(1), 65–81.
- Ciuk, D., Troy, A., & Jones, M. (2015). *Measuring emotion: Self-reports vs. physiological indicators*. 2015 Annual Meeting of the Midwest Political Science Association.
- Clasen, M., Kjeldgaard-Christiansen, J., & Johnson, J. A. (2020). Horror, personality, and threat simulation: A survey on the psychology of scary media. *Evolutionary Behavioral Sciences*, 14(3), 213.
- Codispoti, M., Surcinelli, P., & Baldaro, B. (2008). Watching emotional movies: Affective reactions and gender differences. *International Journal of Psychophysiology*, 69(2), 90–95. <https://doi.org/10.1016/j.ijpsycho.2008.03.004>
- Critchley, H. D., Rotshtein, P., Nagai, Y., O’Doherty, J., Mathias, C. J., & Dolan, R. J. (2005). Activity in the human brain predicting differential heart rate responses to emotional facial expressions. *Neuroimage*, 24(3), 751–762.
- D’Ydewalle, G., & De Bruycker, W. (2007). Eye movements of children and adults while reading television subtitles. *European Psychologist*, 12(3), 196–205. <https://doi.org/10.1027/1016-9040.12.3.196>
- D’Ydewalle, G., & Gielen, I. (1992). Attention allocation with overlapping sound, image, and text. In K. Rayner (Ed.), *Eye movements and visual cognition: Scene perception and reading* (pp. 414–427). Springer-Verlag.
- D’Ydewalle, G., Praet, C., Verfaillie, K., & Van Rensbergen, J. (1991). Watching subtitled television: Automatic reading behavior. *Communication Research*, 18(5), 650–666. <https://doi.org/10.1177/009365091018005005>
- Delabastita, D. (1990). Translation and the mass media. In S. Bassnett & A. Lefevere (Eds.), *Translation, history and culture* (pp. 97–109). The Pinter Press.
- Di Giovanni, E. (2016). Reception studies in audiovisual translation research. The case of subtitling at film festivals. *trans-kom*, 9(1), 58–78. http://www.trans-kom.eu/bd09nr01/trans-kom_09_01_05_Di_Giovanni_Reception.20160705.pdf
- Di Giovanni, E., & Gambier, Y. (2018). *Reception studies and audiovisual translation*. John Benjamins.
- Díaz Cintas, J. (2019). Audiovisual translation. In E. Angelone, M. Ehrensberger-Dow, & G. Massey (Eds.), *The Bloomsbury Companion to language industry studies* (pp. 209–230). Bloomsbury.

- Díaz Cintas, J., & Muñoz Sánchez, P. (2006). Fansubs: Audiovisual translation in an amateur environment. *Journal of Specialised Translation*, 6, 37–52.
https://www.jostrans.org/issue06/art_diaz_munoz.pdf
- Díaz Cintas, J., & Remael, A. (2007). *Audiovisual translation: Subtitling*. St. Jerome Publishing.
- Díaz Cintas, J., & Szarkowska, A. (2020). Experimental research in audiovisual translation – Cognition, reception, production. *Journal of Specialised Translation*, 33, 3–16.
https://jostrans.org/issue33/art_introduction.pdf
- Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psychology*, 5, 781.
- Dienes, Z. (2015). How Bayesian statistics are needed to determine whether mental states are unconscious. In: Overgaard, M. (Ed.), *Behavioural methods in consciousness research* (pp. 199–220). Oxford University Press.,
- Dienes, Z. (2016). How Bayes factors change scientific practice. *Journal of Mathematical Psychology*, 72, 78–89.
- Dienes, Z. (2019). How do I know what my theory predicts? *Advances in Methods and Practices in Psychological Science*, 2(4), 364–377.
- Dienes, Z. (2021). How to use and report Bayesian hypothesis tests. *Psychology of Consciousness: Theory, Research, and Practice*, 8(1), 9–26.
- Dillon, C. (2006). *Emotional responses to immersive media* [Unpublished doctoral dissertation]. University of London.
- Doherty, S., & Kruger, J-L. (2018). A systematic review of the eye tracking measures used in empirical research on subtitling and captioning. In T. Dwyer, C. Perkins, S. Redmond, & J. Sita (Eds.), *Seeing into screens: Eye tracking and the moving image* (pp. 46–64). Bloomsbury.
- Ebert, R. (2006). *The day shift is no better*. <https://www.rogerebert.com/reviews/night-watch-2006>
- Ekman, P. (1984). Expression and the nature of emotion. In K. Rainer Scherer & P. Ekman (Eds.), *Approaches to emotion* (pp. 319–344). Lawrence Erlbaum Associates.
- Ekman, P. (1992). Are there basic emotions? *Psychological Review*, 99(3), 550–553.
<https://doi.org/10.1037/0033-295X.99.3.550>
- Ekman, P. (1999). Basic emotions. *Handbook of Cognition and Emotion*, 98, 45–60.
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, 41(4), 1149–1160.
- Fernández, C., Pascual, J. C., Soler, J., Elices, M., Portella, M. J., & Fernández-Abascal, E. (2012). Physiological responses induced by emotion-eliciting films. *Applied Psychophysiology and Biofeedback*, 37(2), 73–79. <https://doi.org/10.1007/s10484-012-9180-7>
- Foerster, A. (2010). Towards a creative approach in subtitling: a case study. In J. Díaz Cintas, A. Matamala, & J. Neves (Eds.), *New insights into audiovisual translation and media accessibility: Media for all 2* (pp. 81–98). Rodopi.
- Fox, W. (2016). Integrated titles: An improved viewing experience? In S. Hansen-Schirra & S. Grucza (Eds.), *Eyetracking and applied linguistics* (pp. 5–31). Language Science Press.
- Fox, W. (2018). Can integrated titles improve the viewing experience? Investigating the impact of subtitling on the reception and enjoyment of film using eye tracking and questionnaire data. *Translation and Multilingual Natural Language Processing*, 9, Language Science Press.
<https://doi.org/10.5281/zenodo.1180721>
- Fryer, L. (2013). Putting it into words: The impact of visual impairment on perception, experience and presence [Unpublished doctoral dissertation]. University of London.

- Garnefski, N., & Kraaij, V. (2006). Cognitive emotion regulation questionnaire—development of a short 18-item version (CERQ-short). *Personality and Individual Differences*, *41*(6), 1045–1053.
- Goodman, L. A., Corcoran, C., Turner, K., Yuan, N., & Green, B. L. (1998). Assessing traumatic event exposure: General issues and preliminary findings for the Stressful Life Events Screening Questionnaire. *Journal of Traumatic Stress*, *11*(3), 521–542.
- Gottlieb, H. (1994). *Subtitling: Diagonal translation*. *Perspectives*, *2*(1), 101–21.
<https://doi.org/10.1080/0907676X.1994.9961227>
- Gray, M. J., Litz, B. T., Hsu, J. L., & Lombardo, T. W. (2004). Psychometric properties of the life events checklist. *Assessment*, *11*(4), 330–341.
- Grenoble, L. A. (2010). Contact and the development of the Slavic languages. In R. Hickey (Ed.), *The handbook of language contact* (pp. 581–597). Wiley-Blackwell.
- Gross, J. J., & Levenson, R. W. (1995). Emotion elicitation using films. *Cognition & Emotion*, *9*(1), 87–108.
- Hickie, I. B., Davenport, T. A., Hadzi-Paviovic, D., Koschera, A., NCLIPmith, S. L., Scott, E. M., & Wilhelm, K. A. (2001). Development of a simple screening tool for common mental disorders in general practice. *Medical Journal of Australia*, *175*, S10–S17.
- Hong, R., Wang, M., Yuan, X.-T., Xu, M., Jiang, J., Yan, S., & Chua, T.-S. (2011). Video accessibility enhancement for hearing-impaired users. *ACM Transactions on Multimedia Computing, Communications, and Applications*, *7*(1), 1–19. <https://doi.org/10.1145/2037676.2037681>
- Iturregui-Gallardo, G., & Matamala, A. (2020). Audio subtitling: Dubbing and voice-over effects and their impact on user experience. *Perspectives*, *29*, 64–83.
<https://doi.org/10.1080/0907676X.2019.1702065>
- Iturregui-Gallardo, G., & Soler-Vilageliu, O. (2021). Audio subtitling and subtitling: A comparison of their emotional effect on blind / partially sighted and sighted users. *Onomázein*, 61–82.
<https://doi.org/10.7764/onomazein.ne8.07>
- Ivarsson, J., & Carroll, M. (1998). *Subtitling*. TransEdit.
- Izard, C. E. (1977). *Human Emotions*. Plenum.
- Izard, C. E. (1993). Four systems for emotion activation: Cognitive and noncognitive processes. *Psychological Review*, *100*(1), 68–90. <https://doi.org/10.1037/0033-295X.100.1.68>
- Kabara, T. (2015). What is gained in subtitling: How film subtitles can expand the source text. *TransCulturAl: Journal of Translation and Cultural Studies*, *7*(1), 166–179.
- Kramer, M. D., Patrick, C. J., Hettema, J. M., Moore, A. A., Sawyers, C. K., & Yancey, J. R. (2020). Quantifying dispositional fear as threat sensitivity: Development and initial validation of a model-based scale measure. *Assessment*, *27*(3), 533–546.
- Kreibig, S. D. (2010). Autonomic nervous system activity in emotion: A review. *Biological Psychology*, *84* (3), 394–421. <https://doi.org/10.1016/j.biopsycho.2010.03.010>
- Kruschke, J. K., & Liddell, T. M. (2018). The Bayesian new statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. *Psychonomic Bulletin & Review*, *25*, 178–206.
- Kruger, J. L., Doherty, S., Fox, W., & de Lissa, P. (2018). Multimodal measurement of cognitive load during subtitle processing: Same-language subtitles for foreign language viewers. In I. Lacruz & R. Jääskeläinen (Eds.), *Innovation and expansion in translation process research* (pp. 267–294). John Benjamins.
- Kruger, J-L., Hefer, E., & Matthew, G. (2013). Measuring the impact of subtitles on cognitive load: Eyetracking and dynamic audiovisual texts. *Proceedings of the 2013 Conference on Eye Tracking South Africa*, 62–66. <https://dl.acm.org/doi/pdf/10.1145/2509315.2509331>

- Kruger, J.-L., Hefer, E., & Matthew, G. (2014). Attention distribution and cognitive load in a subtitled academic lecture: L1 vs. L2. *Journal of Eye Movement Research*, 7(5), 1–15.
<https://doi.org/10.16910/jemr.7.5.4>
- Kruger, J.L., Soto-Sanfiel, M. T., Doherty, S., & Ibrahim, R. (2016). Towards a cognitive audiovisual translatology: Subtitles and embodied cognition. In R. Muñoz (Ed.), *Reembedding translation process research* (pp. 171–194). John Benjamins.
- Lakens, D., Adolphi, F. G., Albers, C. J., Anvari, F., Apps, M. A., Argamon, S. E., ... & Zwaan, R. A. (2018). Justify your alpha. *Nature Human Behaviour*, 2(3), 168–171.
- Langer Rossi, L. (2019). The translated text: The aesthetic impact of language on transnational cinema. In R. Junkerjürgen & Rebane, G. (Eds.), *Multilingualism in film* (pp. 35–56). Peter Lang.
- Leveridge, F., & Mével, P-A. (2023). Cinéma-monde and subtitling. *Journal of Romance Studies*, 23(1). <https://doi.org/10.3828/jrs.2023.16>
- Liu, C. C., & Aitkin, M. (2008). Bayes factors: Prior sensitivity and model generalizability. *Journal of Mathematical Psychology*, 52(6), 362–375.
- Maj, M. (2008). Depression, bereavement, and “understandable” intense sadness: Should the DSM-IV approach be revised? *The American Journal of Psychiatry*, 165(11), 1373–1375.
<https://doi.org/10.1176/appi.ajp.2008.08071047>
- Mandler, G. (1992). Memory, arousal, and mood: A theoretical integration. In S-Å. Christianson (Ed.), *The handbook of emotion and memory: Research and theory* (pp. 93–110). Erlbaum.
- Marleau, L. (1982). *Les sous-titres... un mal nécessaire* [Subtitles... a necessary evil]. *Meta*, 27(3), 271–285. <https://doi.org/10.7202/003577ar>
- Matamala, A., Soler-Vilageliu, O., Iturregui-Gallardo, G., Jankowska, A., Méndez-Ulrich, J.-L., & Serrano, A. (2020). Electrodermal activity as a measure of emotions in media accessibility research: Methodological considerations. *Journal of Specialised Translation*, 33, 129–151.
https://jostrans.org/issue33/art_matamala.pdf
- McCambridge, J., Witton, J., & Elbourne, D. R. (2014). Systematic review of the Hawthorne effect: New concepts are needed to study research participation effects. *Journal of Clinical Epidemiology*, 67(3), 267–277.
- McClarty, R. (2012). Towards a multidisciplinary approach in creative subtitling. *MonTI*, 4(4), 113–153. <https://doi.org/10.6035/MonTI.2012.4.6>
- McClarty, R. (2013). *Film and translation: The art of subtitling* [Unpublished doctoral dissertation]. Queen’s University Belfast.
- Mével, P-A. (2017). *Subtitling African American English into French: Can we do the right thing?* Peter Lang.
- Mével, P-A. (2020). Accessible paratext: Actively engaging (with) D/deaf audiences. *Punctum*, 6(1), 203–219. <https://punctum.gr/download/accessible-paratext-actively-engaging-with-d-deaf-audiences/?tmstv=1674690885>
- Mével, P-A., Robinson, J., & Tennent, P. (2022). Immersive, creative, inclusive: Areas of cross-fertilization between accessible captions for D/deaf audiences for the stage and for the screen. *Journal of Audiovisual Translation*, 5(2), 176–193.
<https://doi.org/10.47476/jat.v5i2.2022.227>
- Neves, J. (2005). *Audiovisual translation: Subtitling for the deaf and hard-of-hearing* [Unpublished doctoral dissertation]. University of Surrey-Roehampton.

- Nie, D., Wang, X-W., Shi, L-C., & Lu, L-B. (2011, April 27 – May 1). EEG-based emotion recognition during watching movies [Conference presentation], In *2011 5th International IEEE/EMBS Conference on Neural Engineering* (pp. 667–670). Cancun, Mexico.
<https://doi.org/10.1109/NER.2011.5910636>
- Nornes, A. (1999). For an abusive subtitling. *Film Quarterly*, 52(3), 17–34.
<https://doi.org/10.2307/1213822>
- O’Sullivan, C. (2011). *Translating popular film*. Palgrave Macmillan.
- Oliver, M. B., & Sanders, M. (2004). The appeal of horror and suspense. In S. Prince (Ed.), *The horror film* (pp. 242–260). <https://doi.org/10.36019/9780813542577>
- Orero, P., Doherty, S., Kruger, J-L., Matamala, A., Pedersen, J., Perego, E., Romero-Fresco, P., Rovira-Esteva, S., Soler-Vilageliu, O., & Szarkowska, A. (2018). Conducting experimental research in audiovisual translation (AVT): A position paper. *Journal of Specialised Translation*, 30, 105–126. https://www.jostrans.org/issue30/art_orero_et_al.pdf
- Pedersen, J. (2007). Cultural Interchangeability: The effects of substituting cultural references in subtitling. *Perspectives*, 15(1), 30–48. <https://doi.org/10.2167/pst003.0>
- Pedersen, J. (2017). The FAR model: Assessing quality in interlingual subtitling. *Journal of Specialised Translation*, 28, 210–229. https://www.jostrans.org/issue28/art_pedersen.pdf
- Perego, E., Del Missier, F., Porta, M., & Mosconi, M. (2010). The cognitive effectiveness of subtitle processing. *Media Psychology*, 13(3), 243–272.
<https://doi.org/10.1080/15213269.2010.502873>
- Pérez-González, L. (2014). *Audiovisual translation: Theories, methods and issues*. Routledge.
- Petrantonakis, P. C., & Hadjileontiadis, L. J. (2010). Emotion recognition from EEG using higher order crossings. *IEEE transactions on information technology in biomedicine: A publication of the IEEE Engineering in Medicine and Biology Society*, 14(2), 186–197.
<https://doi.org/10.1109/TITB.2009.2034649>
- Petrantonakis, P. C., & Hadjileontiadis, L. J. (2011). A novel emotion elicitation index using frontal brain asymmetry for enhanced EEG-based emotion recognition. *IEEE transactions on information technology in biomedicine: A publication of the IEEE Engineering in Medicine and Biology Society*, 15(5), 737–746. <https://doi.org/10.1109/TITB.2011.2157933>
- Pollard, C. (2019). Applying a multidisciplinary approach to aesthetic subtitling practice: A subtitle commentary on original aesthetic subtitles for “Le Fabuleux Destin d’Amélie Poulain” [Unpublished doctoral dissertation]. University of Geneva.
- Ramos, M. (2015). The emotional experience of films: Soes Audio Description make a difference? *The Translator*, 21(1), 68–94. <https://doi.org/10.1080/13556509.2014.994853>
- Ramos Caro, M. (2016). Testing audio narration: The emotional impact of language in audio description. *Perspectives*, 24(4), 606–634. <https://doi.org/10.1080/0907676X.2015.1120760>
- Romero-Fresco, P., & Fryer, L. (2018). *Accessible filmmaking guide*. Archer’s Mark.
- Romero-Fresco, P. (2019) *Accessible filmmaking: Integrating translation and accessibility into the filmmaking process*. Routledge.
- Sultana, I., Ali, A., & Iftikhar, I. (2021). Effects of horror movies on psychological health of youth. *Global Mass Communication Review*, VI.
- Szarkowska, A. (2005). The power of film translation. *Translation Journal*, 9(2), 1–9.
- Szarkowska, A., & Gerber-Morón, O. (2018). Viewers can keep up with fast subtitles: Evidence from eye movements. *PLoS ONE*, 13(6), e0199331. <https://doi.org/10.1371/journal.pone.0199331>
- Szarkowska, A., & Gerber-Morón, O. (2019). Two or three lines: A mixed-methods study on subtitle processing and preferences. *Perspectives*, 27(1), 144–164.
<https://doi.org/10.1080/0907676X.2018.1520267>

- Taylor S. E. (1991). Asymmetrical effects of positive and negative events: The mobilization-minimization hypothesis. *Psychological Bulletin*, 110(1), 67–85. <https://doi.org/10.1037/0033-2909.110.1.67>
- Thode, H. C. (2002). *Testing for normality*. CRC Press.
- Thompson, J. (2007). Online Alexithymia Questionnaire—G2. <http://oaq.blogspot.com/>
- Tsikandilakis, M., Chapman, P., & Peirce, J. (2018). Target meta-awareness is a necessary condition for physiological responses to masked emotional faces: Evidence from combined skin conductance and heart rate assessment. *Consciousness and Cognition*, 58, 75–89.
- Tsikandilakis, M., Bali, P., Derrfuss, J., & Chapman, P. (2019). The unconscious mind: From classical theoretical controversy to controversial contemporary research and a practical illustration of the “error of our ways”. *Consciousness and Cognition*, 74, 102771.
- Tsikandilakis, M., Bali, P., Derrfuss, J., & Chapman, P. (2020a). “I can see you; I can feel it; and vice-versa”: Consciousness and its relation to emotional physiology. *Cognition and Emotion*, 34(3), 498–510.
- Tsikandilakis, M., Bali, P., Derrfuss, J., & Chapman, P. (2020b). Anger and hostility: Are they different? An analytical exploration of facial-expressive differences, and physiological and facial-emotional responses. *Cognition and Emotion*, 34(3), 581–595.
- Tsikandilakis, M., Leong, M. Q., Yu, Z., Paterakis, G., Bali, P., Derrfuss, J., ... & Mitchell, P. (2021a). “Speak of the Devil... and he Shall Appear”: Religiosity, unconsciousness, and the effects of explicit priming in the misperception of immorality. *Psychological Research*, 1–29.
- Tsikandilakis, M., Yu, Z., Kausel, L., Boncompagni, G., Lanfranco, R. C., Oxner, M., ... & Chapman, P. (2021b). “There Is No (Where a) Face Like Home”: Recognition and appraisal responses to masked facial dialects of emotion in four different national cultures. *Perception*, 50(12), 1027–1055.
- Uhrig, M. K., Trautmann, N., Baumgärtner, U., Treede, R-D., Henrich, F., Hiller, W., & Marschall, S. (2016). Emotion elicitation: A comparison of pictures and films, *Frontiers in Psychology*, 7, 1–12. <https://doi.org/10.3389/fpsyg.2016.00180>
- Van Der Ploeg, M. M., Brosschot, J. F., Versluis, A., & Verkuil, B. (2017). Peripheral physiological responses to subliminally presented negative affective stimuli: A systematic review. *Biological Psychology*, 129, 131–153.
- Van Hoecke, S. M., Schrijver, I., & Robert, I. S. (2022). Methodological preparation of a within-subject audiovisual cognition, reception and perception study. *Journal of Audiovisual Translation*, 5(1), 94–128.
- Wagenmakers, E. J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Love, J., ... & Morey, R. D. (2018a). Bayesian inference for psychology. Part I: Theoretical advantages and practical ramifications. *Psychonomic Bulletin & Review*, 25(1), 35–57.
- Wagenmakers, E. J., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., ... & Morey, R. D. (2018b). Bayesian inference for psychology. Part II: Example applications with JASP. *Psychonomic Bulletin & Review*, 25(1), 58–76.

Appendix

1: SCENE Selections by Group

App. Table 1

Distribution for Hours per Week per SCENE Experience (movies n) Groups

		SCENE Experience (Movies n)	Total Participants	Female Participants
Average Time Movies and Series Watching Per Week (Hours)	One to Four	G1 (0 to 2)	11	7
		G2 (3 to 5)	5	1
		G3 (6 to 8)	0	0
	Five to Eight	G1 (0 to 2)	9	4
		G2 (3 to 5)	4	2
		G3 (6 to 8)	3	2
	Nine to Twelve	G1 (0 to 2)	3	1
		G2 (3 to 5)	8	3
		G3 (6 to 8)	5	4

2: SCENE Selections by Group Stage Two

App. Table 2

Distribution for Hours per Week per SCENE Experience (movies n) Groups

		SCENE Experience (Movies n)	Total Participants	Female Participants
Average Time Movies and Series Watching Per Week (Hours)	One to Four	G1 (0 to 2)	54	26
		G2 (3 to 5)	25	10
		G3 (6 to 8)	11	9
	Five to Eight	G1 (0 to 2)	28	12
		G2 (3 to 5)	43	20
		G3 (6 to 8)	29	25
	Nine to Twelve	G1 (0 to 2)	16	8
		G2 (3 to 5)	33	12
		G3 (6 to 8)	41	25