

Gender characteristics and comparisons in the emotional reception of standard and aesthetically
integrated subtitles in a fantasy-thriller context

Pierre-Alexis Mével ¹

Francesca Leveridge ²

Myron Tsikandilakis ^{1, 3-4}

¹ Department of Modern Languages and Cultures, University of Nottingham

² Department of Modern Languages and Cultures, University of Exeter

³ School of Psychology, University of Nottingham

³ Medical School, Faculty of Medicine and Health Sciences, University of Nottingham

Addresses for correspondence:

Dr Pierre-Alexis Mével
Department of Modern
Languages and Cultures,
University of Nottingham
pierre-alexis.mével@nottingham.ac.uk

Dr Francesca Leveridge
Department of Modern
Languages and Cultures,
University of Exeter
F.Leveridge@exeter.ac.uk

Dr Myron Tsikandilakis
School of Psychology,
School of Medical
and Health Sciences,
Department of Modern
Languages and Cultures,
University of Nottingham
myron.tsikandilakis3@nottingham.ac.uk
tsikandilakismyron@gmail.com

Word Count: 8,034 words

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

Abstract

In this article, we test the findings by Leveridge and colleagues (2024) as regards the reception of aesthetically integrated subtitles (AIS), and most critically we explore the role of gender in these outcomes. We explore psychophysiological (skin conductance and heart rate) and self-report responses (questionnaires) among three different subtitle deliveries (no subtitles, standard subtitles, and aesthetically integrated) overall, and between male and female participants in a series of clips from the Russian thriller-horror film *Night Watch* (Bekmambetov, 2004). In this research, we involve the entire array of film clips that included AIS from the film *Night Watch*. We use null-hypothesis significance-testing (NHST) and Bayesian analyses to explore the claim that AIS enhance psychophysiological arousal and viewing experience quality compared to standard subtitles (STS), and investigate the gender characteristics and differences in the reported outcomes. We provide a replication of previous findings. Critically, we show that female participants exhibited higher arousal for AIS, and also separately for all subtitle types, that was not reflected in significant between-gender differences for self-reports for intensity and arousal. We show, in another novel section of this research, that AIS subtitles increased interest in film genres for both genders, suggesting their significant potential for future audiovisual translation practices.

1. Introduction

The perception of gender disparities in emotional responses stands as one of the most enduring gender stereotypes to exist (Fabes & Martin, 1991; Fischer, 1993; Grossman & Wood, 1993; Plant et al., 2000; Timmers et al., 2003). The notion that women are inherently more emotional than men has in fact been designated a ‘master stereotype’ (Shields, 2003), and one that transcends specific positive and negative emotional responses to discrete elicitors, such as happiness, fear, disgust, and sadness (Shields, 2003; Birnbaum et al., 1980; Briton & Hall, 1995). The impact of these stereotypes and their pervasive societal influence on individuals’ self-reporting of their emotional experiences, that is, reports collected retrospectively via questionnaires, surveys, and interviews, amongst other tools, is significant: on the one hand, it is of course entirely possible that women genuinely experience heightened feelings of fear, anxiety and disgust compared to men, as studies examining responses to aversive audiovisual stimuli through the use of self-reports have suggested (Thunberg & Dimberg, 2000; Codispoti et al., 2008; Rohrmann et al., 2008; Lithari et al., 2010; Bianchin & Angrilli, 2012; Fernández et al., 2012; Maffei et al., 2015; Blanco-Ruiz et al., 2020; Hagemann et al., 1999). However, men’s purportedly subdued responses in these reports may also be rooted in an inclination to conceal emotions that could be perceived to deviate from societal expectations of men to be strong, rational, and able to overcome danger (Plant et al., 2000), even if they feel scared, worried and/or repulsed by the stimuli to which they have been exposed. Whilst self-report measures such as questionnaires and surveys are of course informative, emotional experiences can be subjective. This means that emotions may 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. Nevertheless, as a result, these responses are – statistically – considered ordinal data that do not necessarily involve within- and between-participants agreeable rating characteristics (see Boone, 2020; Bürkner & Vuorre, 2019). This means that when evaluated without physiological assessment technology (Cacioppo et al., 2007) – that is, not controlled for subjectivity, self-presentation and awareness biases (see Ciuk et al., 2015) – they could fail to capture the full array of effects that stimuli induce to participants. Studies measuring the intensity with which emotions are experienced by the strength of visceral arousal, i.e., the activation of components of the autonomic nervous system, such as the heart, and sweat gland activity, also known as heart-rate (HR) and skin conductance responses (SCR) respectively, have provided contrary results concerning whether women are emotionally more sensitive than men (Kring & Gordon, 1998; Bradley et al., 2001; Labouvie-Vief et al., 2003). Some studies using autonomic nervous system activity assessments indicate that there are no gender differences in emotional responses (Levenson et al., 1990; 1991; Kelly et al., 2008). As for the emotional processing of aversive audiovisual stimuli discussed above, and despite leveraging physiological responses, the data is largely inconclusive, owing invariably to issues relating to experimental methodology, participant sample size parameters and gender distribution biases. Further research is thus warranted in this area (see Vrana & Rollock, 2002).

This paper pursues this gender-based line of inquiry into the emotional processing of aversive audiovisual stimuli by extending it to audiovisual translation (AVT). Specifically, it details an experiment carried out by a team of researchers to investigate the emotional correlates of different types of subtitling delivery, such as standard and aesthetically integrated subtitles, on audiences in the context of fear-eliciting clips from the Russian fantasy thriller film *Night Watch* (Bekmambetov, 2004). The experiment employed a methodology combining skin conductance responses (SCR) heart rate responses (HR) (see *Methods: Psychophysiology*) and

self-reports (questionnaires) to account for the complex interplay between experiential, cognitive, behavioural, and physiological elements that make up emotional responses, null-hypothesis-significance-testing (NHST) and Bayesian analyses were employed to compare the subtitling deliveries (see *Methods: Statistical Framework*). A previous iteration of this experiment (Leveridge et al., 2024) demonstrated that the presentation of aesthetically integrated subtitles (AIS; see section 2.2 below) led to higher positively rated psychophysiological arousal and quality of viewing experience ratings compared to standard subtitles. The effect of gender was not explored in physiological or self-report responses in that manuscript. Here we explore whether the novel findings from Leveridge and colleagues (2024) can be replicated and explore the effect of gender in the same context. The specifics of the experiment are described below, followed by a discussion of our findings. We first review the literature on subtitling reception and gender, and on the reception of experimental AVT practices, with particular reference to AIS and psychophysiology.

1. Literature Review

2.1 Gender and Subtitling Reception

Research into the gender dynamics of subtitling reception is notably sparse, with the few existing outcomes inevitably influenced by the intersectionality of gender with other sociocultural characteristics. For example, both younger men and younger women in Italy identified subtitling as their preferred mode of audiovisual translation (see Di Giovanni, 2012). In a Spanish context, young participants reported enjoying amateur subtitling practices, while their older male and female counterparts did not, inarguably revealing to us more about age than gender and the evolving landscape of AVT and its consumption in traditionally dubbing-centric countries (see Orrego-Carmona, 2016)¹. Elsewhere, Gürkan (2019) proposed that higher education levels among female participants in a Turkish study on subtitling for the d/Deaf and hard-of-hearing (SDH) explained their inclination towards strategies demanding greater cognitive effort, such as the use of labels for speaker identification and paralinguistic information. Gürkan suggested that the lower education levels of male participants correlated with their dislike for these same strategies.

Reception study data have also been harnessed by Bosseaux (2023) as a means to tailor the translation of gendered audiovisual narratives, and more specifically in a multilingual documentary about domestic violence. Bosseaux solicited feedback from focus groups on the translation modes – subtitling and dubbing – used to translate the documentary, screened to participants before the discussion, and explored which of the techniques and strategies used therein did the most justice to the voices of the survivors who featured in it. While Bosseaux’s study undoubtedly focused on the *representation* of gender in audiovisual translation, a common emphasis in gender-aware contributions to AVT scholarship (see von Flotow & Josephy-Hernández, 2018), the integration of a *reception* element invaluablely allows us to move beyond speculation about audience sentiments and responses to strategies for translating gendered content and finding out how they really feel. Interestingly, in the ‘Experimental Guidelines for Translating Emotional Content in Documentaries’ (2023) produced as a result of the study, Bosseaux hinted towards the affective potential of certain subtitling strategies, such as that ‘the placement of subtitles is paramount for rendering emotion’ (see Bosseaux, 2023, p. 9). Nevertheless, it is unclear which emotions Bosseaux is referring to, and what their actual or suggested impact on audiences are, or may be, thus opening up avenues for further research. As an example, Blanco-Ruiz and colleagues (2020) carried out a study which relied on self-reports to examine whether audiences

¹ See also Ameri and colleagues (2018) for a similar study conducted in Iran.

experienced heightened fear when watching content depicting violence against women. They found that values reported by both genders are not different for positive emotions but tend to differ for negative emotions. Such studies could be contrasted with the physiological signals of the participants, using HR or SCR. In addition, further research could explore the presence/absence and features (font, size, colour) of subtitles that accompany the clips used in an experiment, or do not, via the employment of subtitles in different fonts and colours, sizes – visual elements that Bosseaux (2023) argues must be considered when translating emotional content.

2.2 Experimentation in AVT and its Reception

Bosseaux's work is particularly relevant to the present manuscript given its focus on gender, emotion, and the use of AIS. the latter being one of several terms to have emerged in recent decades to refer to what Bosseaux calls 'experimental' subtitling practices. This is also referred to as 'creative subtitles' (McClarty, 2012), 'dynamic subtitles' (Brown et al., 2015), 'integrated titles' (Fox, 2016; 2018) and 'free-form subtitles' (Bassnett et al., 2022). AIS refers to subtitles that experiment with typeface, font size, placement, colours, and special effects to correspond closely to the visual style and action of a given audiovisual work. In so doing, AIS override 'standard' subtitling conventions (such as those found in the *Code of Good Subtitling Practice* (Ivarsson & Carroll, 1998) and have been suggested to circumvent issues such as the use of plain white or yellow fonts and the fixing of subtitles to the bottom of the screen covering up, clashing with and/or being concealed by plot-relevant images and objects (see for example Fox, 2018; p. 69; Romero-Fresco, 2019; p. 93). Standard subtitling practices can cause the viewer to miss important pieces of information and hinder their overall comprehension of the narrative, as well as lead to loss of aesthetic value. Whilst we refrain here from delving into epistemological debates about the inherent creativity, or assumed lack thereof, of both types of subtitling practices that have been raised in the process of trying to name them (see Bassnett et al., 2022; Leveridge et al., 2024), we maintain that there are clear differences between subtitles designed to explicitly support the diegetic context and narrative of a film through intentional aesthetic interplay with other semiotic systems, and those that do not. Subtitles – whether integrated or not – may be added to a film post-hoc, as was the case for the aesthetically integrated subtitles of *Night Watch* that were created with English-speaking members of the audiences in mind. The subtitles may also be part of the original production, like in the *John Wick* series for instance, and or in turn translated for foreign-language audiences. For *Night Watch* and for the *John Wick* series, the directors (Bekmambetov and Stahelski respectively) had a direct influence on the creation of the subtitles, which confers potentially high context-value and context-punctuation authority to the subtitling delivery (see Rawsthorn, 2007; Graham, 2014).

There is strong evidence, gathered by means of experimental reception studies, to suggest that giving subtitlers 'more freedom to create an aesthetic that matches that of the text' (McClarty 2012; p. 140) than is typically permitted by industry standards, allows for audiences to fully benefit from these creative functions. Numerous eye-tracking studies conducted on audience reception to AIS (see Caffey, 2009; Brooks & Armstrong, 2014; Brown et al., 2015; Fox, 2016; Black, 2020), as well as one using electroencephalography (EEG; Kruger et al., 2018), suggested that the integration of the subtitles into the aesthetics and visual style of a film reduced the time and effort spent on reading, and increased that spent on the scenes. The concepts discussed here also present striking parallels to those mentioned above by Leveridge and colleagues (2024): AIS not only show higher psychophysiological responses in audiences than standard subtitles, but standard subtitles also elicit lesser psychophysiological effects than when the film clips are viewed with no subtitles at all. The claim that integrated subtitles bridge the gap between the experience of the viewers of an original

work and those of its translated and/or accessible versions is thus well supported (Romero-Fresco & Fryer 2018; p. 13).

The emotional reception of subtitled films has received very little attention – not least on an experimental basis. Experimental studies of this nature, for their part, have found more significant currency in the field of media accessibility (MA). Psychophysiological experiments by Iturregui-Gallardo and Matamala (2020), and Iturregui-Gallardo and Soler-Vilageliu (2021) have focussed on content made accessible via audio subtitling (AST). This is a service that allows for people with visual or reading impairments to access written subtitles in their aural form. Fryer (2013), Ramos (2015) and Ramos Caro (2016) have explored the audio description (AD) of different delivery types: human, synthetic, objective and subjective. Each of these studies employed HR and SCR alongside self-reports to examine reactions to combinations of fear or sadness or disgust, and neutral emotions. Whilst the self-reported measures presented statistically significant differences, particularly when it came to fear, the results from the physiological assessments were mostly inconsistent. As previously shown in Leveridge and colleagues (2024), this issue may be rectified by concentrating on one emotion. Therefore, in the current manuscript we chose to focus on fear because it is, a) one of six basic or universal emotions (Ekman, 1999) alongside anger, surprise, disgust, sadness and happiness, and b) one of the most widely explored emotions in linguistics and psychophysiology as regards the anticipated correlates of arousal it elicits (Kreibig, 2010). It is, therefore, more experimentally well-explored as to how to induce and measure (Taylor, 1991; Uhrig et al., 2016), and it is an important rallying point for applying quantitative methods when it comes to exploring the emotional correlates of creative subtitles.

2. Hypotheses

Our objectives for the current research were to assess the psychophysiological (SCR & HR) and self-report responses among three different subtitle deliveries (no subtitles, standard subtitles, and AIS) overall, and between male and female participants in a series of clips from the Russian thriller-horror film *Night Watch* (Bekmambetov, 2004). Our experimental hypotheses were that AIS would overall provide outcomes for higher positively rated physiological arousal, and that no subtitling would replicate the intriguing finding in Leveridge and colleagues (2024) and provide higher positively rated physiological arousal compared to standard subtitling. Critically, without a-priori hypotheses – given the diversity of previous outcomes in topical research – we explored thoroughly and in statistical detail whether, in what way and to what extent male and female participants differ in their viewer characteristics, and psychophysiological and self-report questionnaire responses to each subtitling delivery in the presented clips.

3. Experiment

Participants: A power calculation based on a between-subjects design and small effect sizes ($f = .1$; $d = .2$; Faul et al., 2009) given the experimental trial contour ($k = 22$; Baker et al., 2020) revealed that ninety-six participants were required for $P_{(1-\beta)} \geq .9$; ($p \leq .05$; $P(H1) \geq .9$; $B > 3$; $d [\geq .2, +\infty]$). One hundred and seven volunteers were screened. The exclusion criteria were that participants should not have seen the film *Night Watch* or seen short scenes, trailer advertisements and film-related commentaries of the film in any format. The inclusion criteria were that participants should currently identify with their gender of birth irrespective and without assessments of their sexual preferences (see Mayer & McHugh, 2016; see also Sun et al., 2019). Participants should have been raised speaking the English language by monolingual non-dyslexic English-speaking parents. The participants should be monolingual non-dyslexic English speakers who completed or

undertook their studies in the English language and were not taught or self-educated in Russian or a related language of Slavic origin.

The participants were invited in an initial screening session lasting approximately one hour. They were screened for clinical and subclinical conditions. They should not have a current or previous DSM-5 Axis I or II diagnosis (American Psychiatric Association, 2019) or a diagnosis for dyslexia; self-report (Y/N). To confirm the self-reports, they were screened with the Somatic and Psychological Health Report Questionnaire (SPHRQ; Berryman et al., 2012), with the Stressful Life Events Screening Questionnaire (SLESQ; Allen et al., 2015) and the Behavior Rating Inventory of Executive Functions 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 Alexithymia, 2024).

The participants were assessed with general emotional sensitivity and sensitivity to fear questionnaires. They were assessed with the Cognitive Emotional regulation Questionnaire (CERQ-short; Garnefski & Kraaij, 2006) and the Trait Fear Scale (TFS; Kramer et al., 2020). The participants were asked to rate their familiarity with subtitles in general (GSF) from one (not at all) to nine (very much), to rate their familiarity with AIS (AIS-F), to choose from a list of commercially available films featuring AIS, series and animations those that they have watched (AIS-A), whether they have watched each selected item more than once (AIS-B), and how many times they have watched each selected item to the best of their recollection (AIS-C). After the completion of the aforementioned assessments, participants were also asked to rate from one (not at all) to nine (very) in a multiple Likert-scales engagement task how much they would be interested to see films and TV shows from a list including romantic/romance, comedy, mystery, drama, action, thriller/horror, animated (genre of preference), historical and documentary, other (genre of preference) and overall (self-report) irrespective of genre.

To balance the final population sample, we used a semi-Bayesian sampling method (see Tsikandilakis et al., 2021). Male and female participants were controlled for assessments that did not relate to gender-related differences (SPHRQ; SLESQ; BRIEF-P/T) and were not controlled for variations that have been suggested to show gender-related differences (CERQ-short; TFS; see Plevkova et al., 2020; but see also Handley et al., 2015). Data from a total of eleven participants were excluded from further analyses due to potential psychiatric conditions (SPHRQ: $n = 5$), potential Alexithymia traits (AQ: $n = 4$) and extreme outlier scores (± 1.5 interquartile range) for life-stressor events (SLESQ: $n = 2$). The final population sample consisted of ninety-six participants (forty-eight females; see Table 1).

Table 1: Participant Characteristics for the Final Population Sample

Participants	n	Age Mean (SD)	B-P/T Mean (SD)	SLESQ Mean (SD)	CERQ-short Mean (SD)	FTQ Mean (SD)	GSF Mean (SD)	AIS-F Mean (SD)	AIS-A Median (Range)	
					CR	ES	TFS			
Overall	96	28.14 (3.14)	46.41 (5.51)	1.51 (.52)	28.97 (7.76)	8.47 (1.16)	2.11 (.84)	5.16 (.9)	5.49 (1.02)	4 (0 – 9)
Male	48	28.19 (3.11)	46.19 (5.44)	1.54 (.52)	29.22 (7.27)	8.47 (.87)	2.14 (.86)	5.19 (.93)	6.81 (1.24)	5 (1 – 9)
Female	48	28.11 (3.17)	46.62 (5.57)	1.47 (.51)	28.72 (6.26)	8.34 (.84)	2.09 (.82)	5.13 (.88)	4.18 (.79)	3 (0 – 8)
NHT and Bayesian Analyses for Gender Comparisons										
P-value (Cohen’s d)			.71 (.02)	.54 (.03)	.72 (.07)	.47 (.14)	.72 (.06)	.73 (.05)	> .001*	= .005*
Bayes Factor			.06	.06	.12	.24	.12	.09	+ ∞	+ ∞

Table 1: This table includes the demographics for a semi-Bayesian sampling for male and female participants. Male and female participants were matched for assessments that did not relate to gender characteristics, such as the SPHRQ, SLESQ and BRIEF-P/T, and were not controlled for variations that could relate to gender-specific differences, such as assessments for CERQ-short and TFS. Asterisks (*) show significant differences between genders for a designated category. For convenience we briefly recapitulate here (see *Procedures*) that B-P/T refers to an assessment for Dyslexia, SLESQ to an assessment for life-stressor events, CERQ to assessments for emotional regulation, FTW/TFS to assessments for fear sensitivity, and GSF to an assessment for subtitling familiarity, AIS-F to an assessment for familiarity with AIS and AIS-A to the number of films with AIS that participants had watched prior to the experiment.

Figure 1: Genre of Preference Film-Interest Ratings per Gender

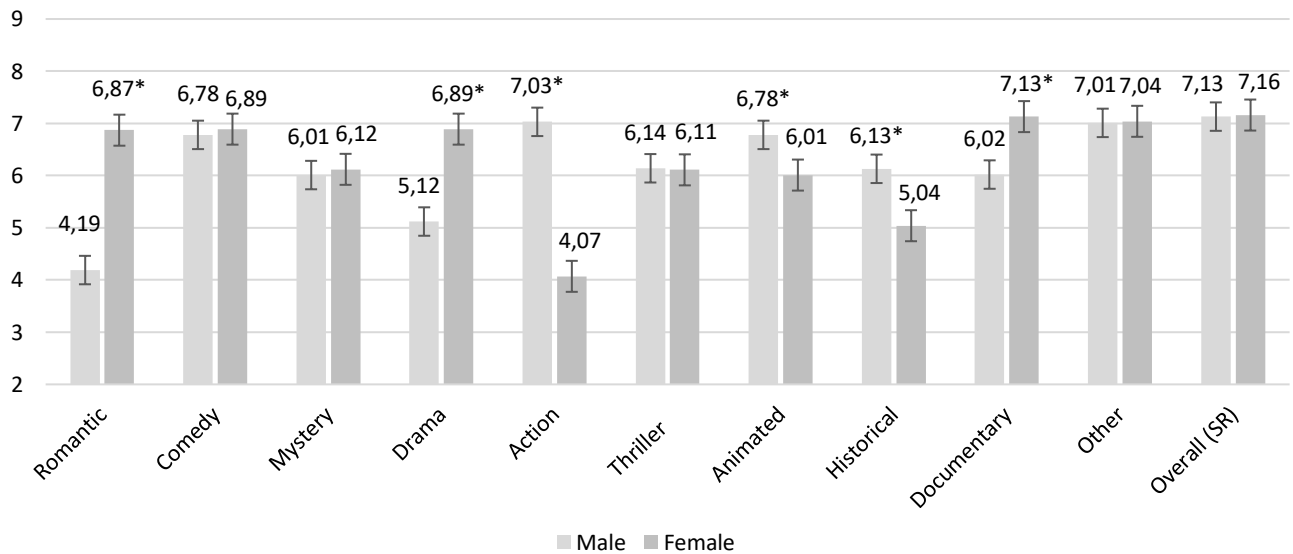


Figure 1: Ratings (one to nine) of genre of preference per gender. Asterisks (*) show significant differences at $p < .001$. the most common items in category "Other" were for males western (24%), science-fiction (21%), fantasy (19%), martial art (19%), and reality show (11%). The most common items for females were fantasy (31%), reality show (25%), reality-show competition (21%) and musical (18%) (for full details of task demographics, see Wuhr et al., 2017; see particularly, Figure 5AC).

Stimuli: The audiovisual stimuli in the main experiment included the twenty-two clips that involved AIS in the fantasy-thriller film *Night Watch*. These 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 duration ranged from fifteen to seventy-nine seconds. Their mean duration was 41.23 seconds (SD = 19.63). Audially, the clips were presented in the Russian language. Their subtitling delivery involved three different formats:

(A.) with English AIS, as included in the cinema version of the film, (B.) with non-AIS/standard English subtitles (STS), as included in the DVD version of the film released in the UK, and (C.) without any subtitles of any kind.

Procedures: The participants were invited in a quiet laboratory space. To avoid external disturbances, noise-cancelling full-ear headphones were provided to the participants during the experiment. The participants were briefed concerning the terminology of the experiment and spent a five-minute on-screen session becoming acquainted with the mouse and keyboard response components of the experimental procedure. Subsequently, they were asked whether they understood the experimental response process (Y/N; self-report) and were given the opportunity to ask the attending researcher any questions relating to the experiment; no instances of negative responses or required feedback reported. The experiment involved three phases with a five-minutes break between each phase. In each phase, the experiment started with a fixation cross in the centre of the screen for 3 (± 1) seconds. After the fixation cross, an audiovisual clip was presented. In each phase, a total of twenty-two different clips were presented once with a single type of subtitling each (A. or B. or C.) throughout the three phases. The order of the phases was randomised, and the clip presentations sequence and subtitling delivery were also randomised in each phase. After each presented clip, a five-second blank screen interval was presented. Then 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 affected 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 question sets and the questions within each set were presented in randomised order in each trial. After the completion of all three phases, the participants were given the opportunity for an additional five-minute break. Participants were then asked to rate in a multiple binary-choices and Likert-scales on-screen engagement task whether (Y/N) and how suitable from one (not at all) to nine (very) they believed AIS compared to STS would be as a subtitling method for romantic/romance, comedy, mystery, drama, action, thriller/horror, animated (genre of preference), historical and documentary films and TV shows, other (genre of preference) and overall, for films and TV shows irrespective of genre. They were also asked in a multiple Likert-scales on-screen engagement task how often from one (not at all) to nine (for the entire film or TV show) they believed that AIS should be used for the same list of genres for films and TV shows and overall (self-report), irrespective of genre. The participants were asked how interested they would be to see a film or TV show with AIS from the same list and overall, irrespective of genre. The placement of each genre in the multiple-choice engagement tasks was randomised for each participant. The order of the presentation of the two engagement tasks was also randomised for each participant (see Figure 2). A blank screen was presented after each experimental sequence for seven seconds, to allow physiological responses to return to baseline (see Cacioppo et al., 2007, pp. 164-167).

Figure 2: Experimental Sequence

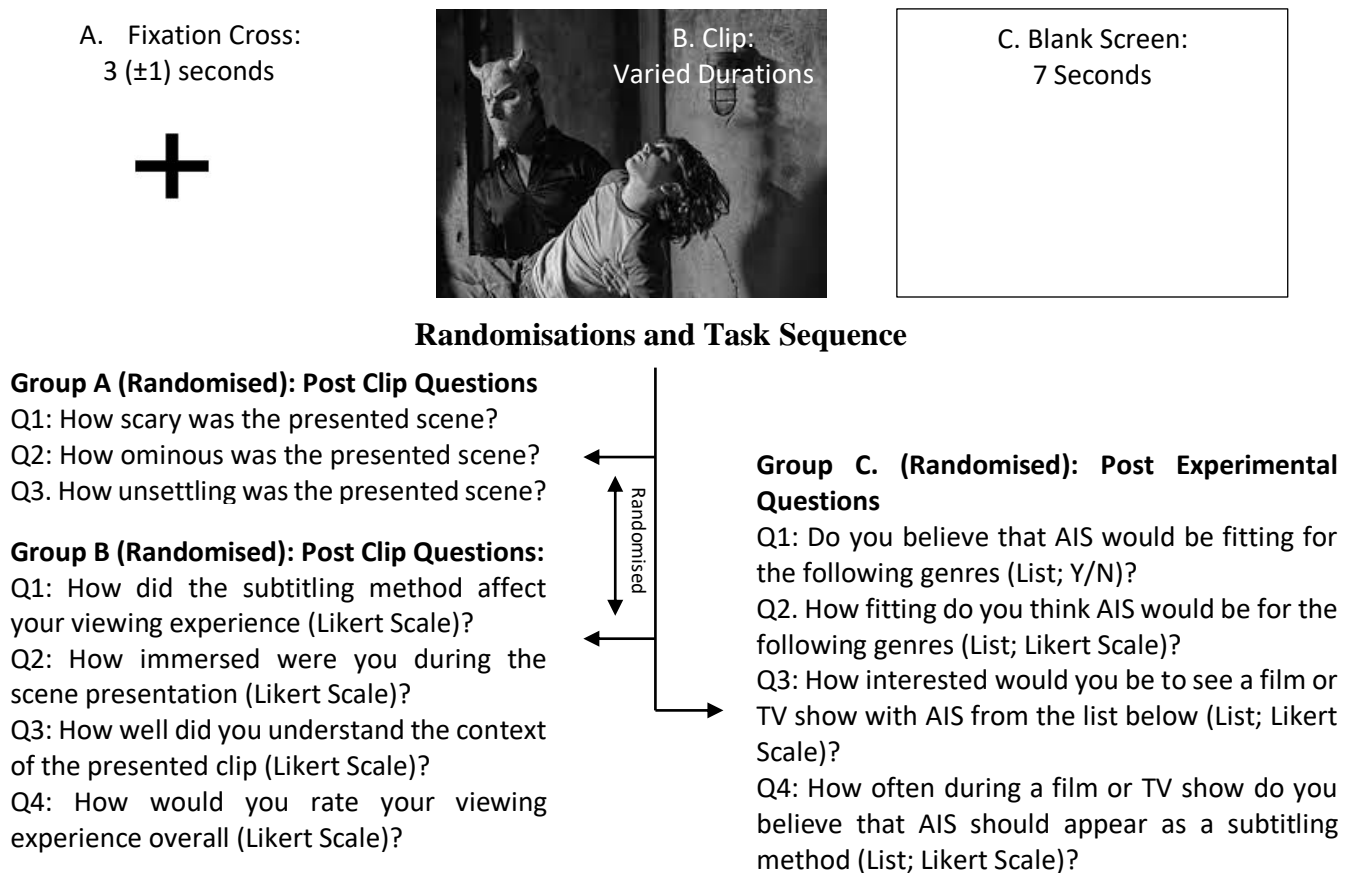


Figure 2: Experimental sequence (for B. with AIS or STS or no subtitles) and participant engagement tasks. Participants were asked to rate after each presentation questions from group A and B; with order of group questions randomised and order of group sequence of questions randomised. After the experiment they were presented with multiple binary-choice and Likert scales (see *Procedures*) with order of questions and Likert scale placement randomised.

Statistical Framework: The analyses were performed using NHST and Bayesian inference. NHST was set at $p \leq .05$ for omnibus ANOVA tests with full post-hoc comparisons including Bonferroni corrections ($\frac{\alpha}{K}$). Bayesian analyses was based on the minimum effect sizes of interest paradigm (Dienes, 2021). We defined evidence for the null hypothesis at a Bayes Factor (BF) $< .33$ given an omnibus effect size of BF (H_0) for Cohen's $f = [0, \leq .05]$. Evidence that the data could be observed under the alternate hypothesis were defined at $B > 3$ for a minimum effect size of interest of BF (H_1) for $f = [\geq .1, +\infty]$, and data that signified inconclusive results at $.33 \leq BF \leq 3$ at BF for $f = [> .05, < .1]$ (for an overview, see Tsikandilakis et al., 2023).

Psychophysiology: For physiological assessment, SCR and HR were used (see Cacciopo et al., 2007; pp. 161-164 & 186-189). Skin conductance was 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. Heart rate was measured via a single-finger sensor from the left hand (ring/third finger). The signal was measured by a BIOPAC System, PPG100C using infrared photoplethysmography and recorded in beats per minute (bpm) in AcqKnowledge. The presence of a phasic skin conductance response was defined as an unambiguous increase ($\mu S \geq .01$) occurring up to three seconds post-stimuli offset. The presence of a heart-rate response was defined as an event-related heart-rate peak in beats per minute occurring up to five seconds post-stimulus offset. Each response was

calculated using the inbuilt derive phasic from tonic and find cycles routines as the highest peak in physiological responses (δ) in respect to a tonic baseline averaged across the period (δT) of each pre-stimulus onset using parallel port-input derived onset markers (for an overview, see Tsikandilakis et al., 2020).

We followed Leveridge and colleagues' (2024) measurement method to provide comparable results. We acquired two different measurements for every presented clip. Firstly, we calculated the tonic SCR and HR baseline during each scene and compared that among AIS, STS and no-subtitle clips to measure the overall physiological changes during these three conditions (see Cacioppo et al., 2007, pp. 164). Secondly, we explored the clips for parallel-input-derived phasic event-related AIS, STS and no-subtitles peak instances, we 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). We acquired for these assessments phasic event-related responses conservatively as unambiguous increases ($SCR \geq \pm .01 \mu S$; $HR \geq \pm 1$ bpm; based on at least three subsequent event-related heart beats; see Cacioppo et al., 2007, p. 188) with reference to the tonic baseline of AIS, STS and no-subtitle scenes (see Cacioppo et al., 2007, pp. 168-173). The raw signals were acquired using the Derive Phasic from Tonic and manual Dirac's delta (δ) functions in AcqKnowledge (see Balakrishnan, 2003) without reported incidents for requirements for additional smoothing, filtering or transformations of the data (see Braithwaite et al., 2013, p. 10-12). Non-responders for physiological changes were included in the analyses (see van der Ploeg et al., 2017; pp. 143-147; see Cacioppo et al., 2007; pp. 164-167 & 187-191; see also Tsikandilakis et al., 2018; 2019, 2020).

Results; Self-Report Ratings: To explore whether we could replicate the Leveridge and colleagues (2024) findings, and explore the impact of gender in the outcomes of the analyses, we conducted a mixed-model multivariate analysis of variance (see Keselman et al., 1998), and Bayesian (BANOVA; see Dong & Wedel, 2014) mixed-model multivariate analyses. The independent variables were Subtitle Delivery (AIS vs STS vs NS) with Gender (Male vs Female) as a between-subjects independent variable. The dependent variables were Self-Report Ratings Responses (Ominous, Scary, Unsettling, Immersion, Context Understanding, Viewing-Experience for Subtitling Delivery, and Viewing Experience Overall; see Procedures above). We reported a significant effect of Subtitle Delivery for all dependent variables ($F(1.82, 170.99) = 978.03$; $p < .001$; $\eta^2_p = .91$; $\epsilon = .76$; Greenhouse-Geiser Corrected; $SE = .008$; $BF = +\infty$) suggesting that we were able to replicate the effects of Leveridge and colleagues (2024) that without exception for all assessed dependent variables, participants rated AIS highest, and NS higher than STS (see Figure 3A.). We could not report any effect and we reported strong evidence for the null for Gender ($F(1,95) = .18$; $p = .92$; $\eta^2_p = .002$; $SE = .012$; $BF = .09$) and a Subtitle-Delivery-by-Gender interaction ($F(1.819, 125.01) = .58$; $p = .56$; $\eta^2_p = .006$; $\epsilon = .77$; Greenhouse-Geiser Corrected; $SE = .017$; $BF = .19$; see Figure 3B.). These findings suggest that we were able to replicate for self-report ratings the findings by Leveridge and colleagues (2024) that AIS were the most preferable subtitling delivery method, and that NS were preferable to STS. Critically, we did not find any gender differences in the participants' responses signifying that both male and female participants provided self-report rating evidence for the replication effects, and that there were no substantial differences and there were evidence for statistical proximity for the reported subtitling delivery preferences between genders (for the reader's convenience, extended results are presented in table and graphic format in Figure 3).

Figure 3: Self-Report Ratings

A. Subtitling Delivery Ratings for Overall Sample							
	Mean (SD)						
	Ominous	Scary	Unsettling	Immersion	Context Underst.	Viewing Experience Subtiting	Viewing Experience Overall
Aesthetically integrated Subtitles (AIS)	6.85 (.33)**	6.86 (.36)**	7.05 (.32)**	6.82 (.27)**	6.93 (.22)**	6.82 (.43)**	7.37 (.26)**
Standard Subtitles (STS)	5.94 (.3)	6.02 (.29)	6.12 (.34)	6.15 (.26)	6.34 (.32)	6.25 (.28)	6.33 (.31)
No Subtitling (NS)	6.19 (.31)*	6.28 (.28)*	6.29 (.32)*	6.44 (.35)*	6.73 (.48)*	6.45 (.35)*	6.78 (.28)*
NHST and Bayesian (BF) Analyses							
p-values (η^2_p)	p < .001 (.46)	p < .001 (.43)	p < .001 (.4)	p < .001 (.38)	p < .001 (.33)	p < .001 (.21)	p < .001 (.55)
Std. Cohen's d (BF)	1.85 (+∞)	1.75 (+∞)	1.64 (+∞)	1.58 (+∞)	1.41 (+∞)	1.03 (+∞)	2.21 (+∞)

B. Subtitling Delivery Ratings per Gender

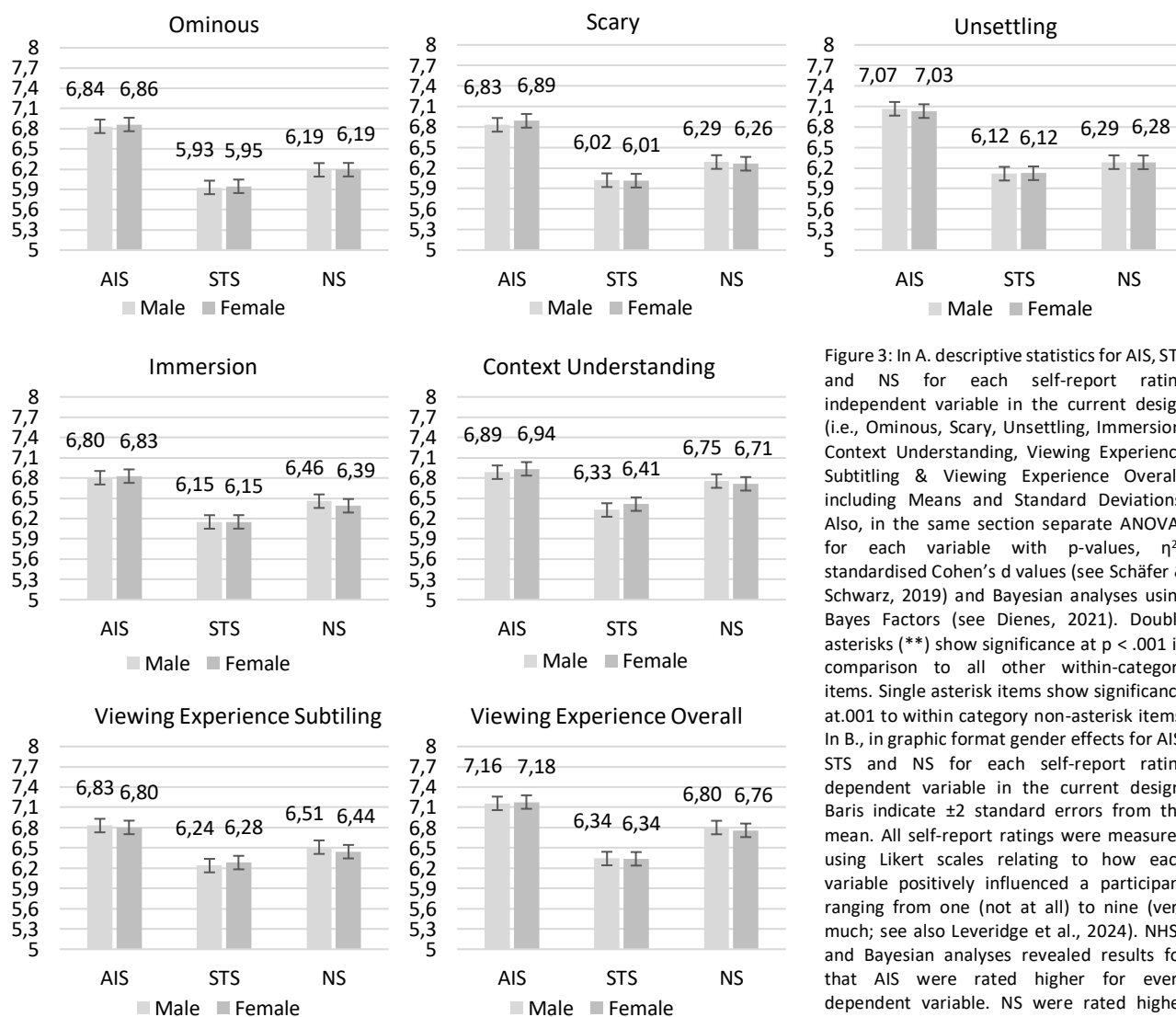


Figure 3: In A. descriptive statistics for AIS, STS and NS for each self-report rating independent variable in the current design (i.e., Ominous, Scary, Unsettling, Immersion, Context Understanding, Viewing Experience Subtiting & Viewing Experience Overall) including Means and Standard Deviations. Also, in the same section separate ANOVAs for each variable with p-values, η^2_p , standardised Cohen's d values (see Schäfer & Schwarz, 2019) and Bayesian analyses using Bayes Factors (see Dienes, 2021). Double asterisks (**) show significance at $p < .001$ in comparison to all other within-category items. Single asterisk items show significance at .001 to within category non-asterisk items. In B., in graphic format gender effects for AIS, STS and NS for each self-report rating dependent variable in the current design. Bars indicate ± 2 standard errors from the mean. All self-report ratings were measured using Likert scales relating to how each variable positively influenced a participant ranging from one (not at all) to nine (very much; see also Leveridge et al., 2024). NHST and Bayesian analyses revealed results for that AIS were rated higher for every dependent variable. NS were rated higher than STS. No effects of gender were reported (see <https://osf.io/x2bp5/>).

Results; Psychophysiology: To explore whether we could replicate the Leveridge and colleagues (2024) findings, and explore the impact of gender in the outcomes of the analyses, we used the same statistical methodology for Psychophysiology as we did for Self-Report Ratings (see Chueh et al., 2012). The independent variables were Subtitle Delivery (AIS vs STS vs NS) with Gender (Male vs Female) as a between-subjects independent variable. The dependent variables were Tonic and Phasic SCR, and HR responses. We reported a significant effect of Subtitle Delivery for all dependent variables ($F(1.79, 168.31) = 88.03$; $p < .001$; $\eta^2_p = .48$; $\epsilon = .89$; Greenhouse-Geiser Corrected; $SE = .163$; $BF = +\infty$). Participants experienced higher arousal to AIS compared to the other subtitling methods, and higher arousal to NS than STS (see Figure Section 4A.). We reported a significant effect of Gender ($F(1,95) = 177.63$; $p < .001$; $\eta^2_p = .32$; $\epsilon = .79$; Greenhouse-Geiser Corrected; $SE = .068$; $BF = +\infty$) and a Subtitle-Delivery-by-Gender interaction ($F(1.84, 173.19) = 43.7$; $p < .001$; $\eta^2_p = .32$; $\epsilon = .79$; Greenhouse-Geiser Corrected; $SE = .068$; $BF = +\infty$) suggesting that although both males and females had similar arousal patterns to subtitling deliveries, females experienced higher arousal for all subtitling deliveries and in all physiological assessment. This provided us a dissonant finding for gender effects between self-reports and psychophysiological measures that we conceptually address thoroughly in the Discussion (for the reader's convenience, extended results for the reported effects are presented in table and graphic format in Figure 4).

Figure 4: Psychophysiology

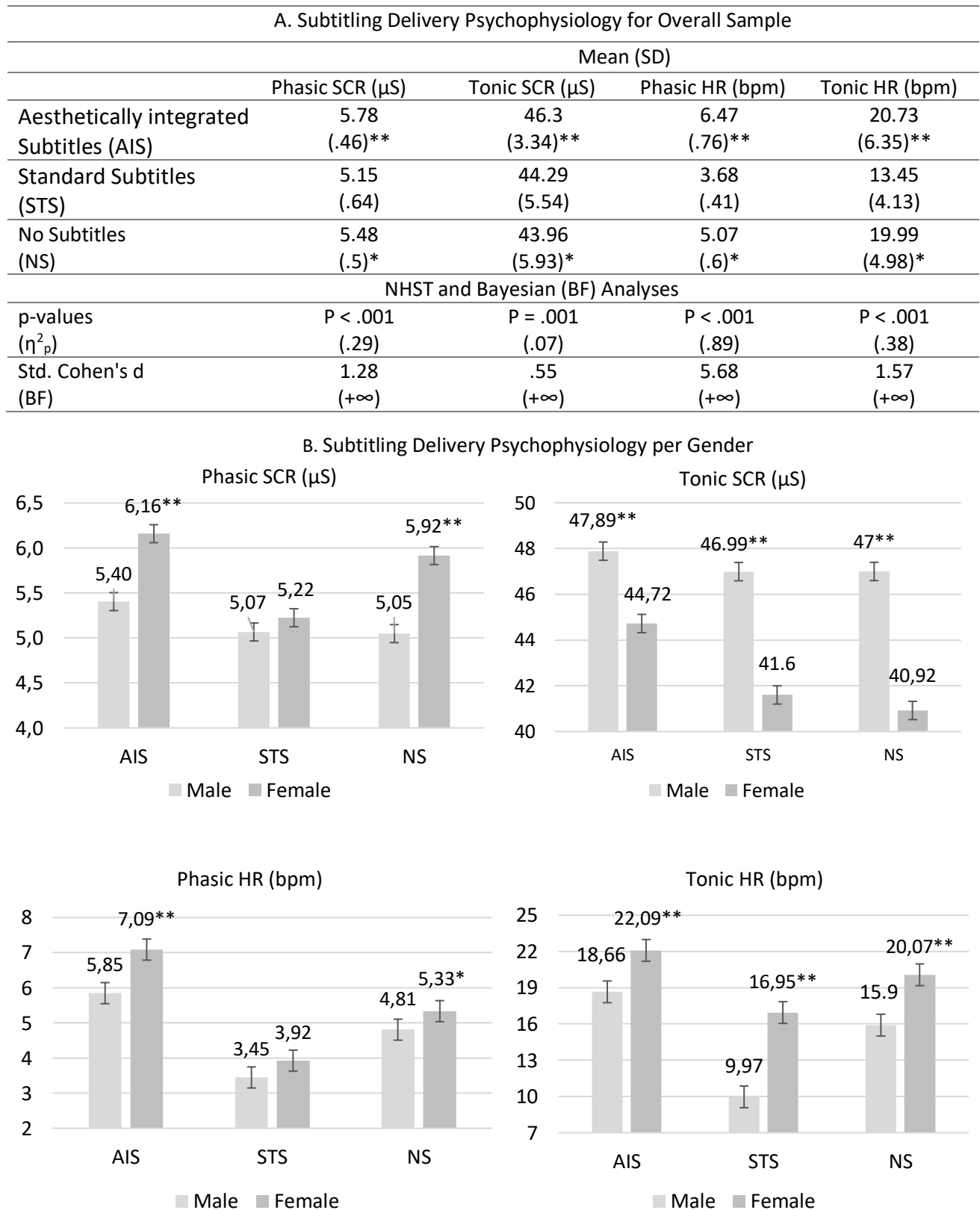


Figure 4: In A. descriptive statistics for AIS, STS and NS for psychophysiological assessments for the overall population, including Means and Standard Deviations. Also, in the same section separate ANOVAs for each variable with p-values, η^2_p , standardised Cohen's d and Bayesian analyses. In B. Gender comparisons per subtitling delivery between male and female participants. In both sections, phasic responses refer to overall arousal from baseline during the entire length of the presented clip. Tonic responses refer to peak arousal incidences during the presentation of each clip for skin conductance and heart rate changes. For both sections, double asterisks (**) show significance at $p < .001$ in comparison to all other within-category items. Single asterisk (*) items show significance at $p < .001$ in comparison to non-asterisk items. In B. bars indicate ± 2 standard errors from the mean.

Results; Post-Experimental Questions: As an exploratory analysis, we compared the gender responses in the experimental questions presented after the completion of the experiment. All participants for both genders replied positively that AIS could be used further in other genres ($M = 100\%$; $SD = 0$). Concerning the questions “How fitting do you think AIS would be for the following genres?”, “How interested would you be to see a film or TV show with AIS from the list below?” and “How often during a film or TV show do you believe that AIS should appear as a subtitling method?” (see Procedures) we did not report significant differences between genres ($p < .05$), and we reported Bayesian evidence for the null ($BF < .33$). The same pattern of results was recorded for differences among genres with the single exception of “How fitting do you think AIS would be for the following genres?” ($F(10, 950) = 80.48$; $p < .001$; $\eta^2_p = .46$; $SE = .01$; $BF = +\infty$). These results suggest that both genders showed similar audience characteristics for ratings for these questions and, furthermore, that, for AIS-including film genres, they did not show the differences reported during the pre-experimental question concerning genre preference-interest illustrated in Figure 1 (for the reader’s convenience, extended results are presented in Figure 5).

Figure 5: Post-experimental AIS Rating Tasks

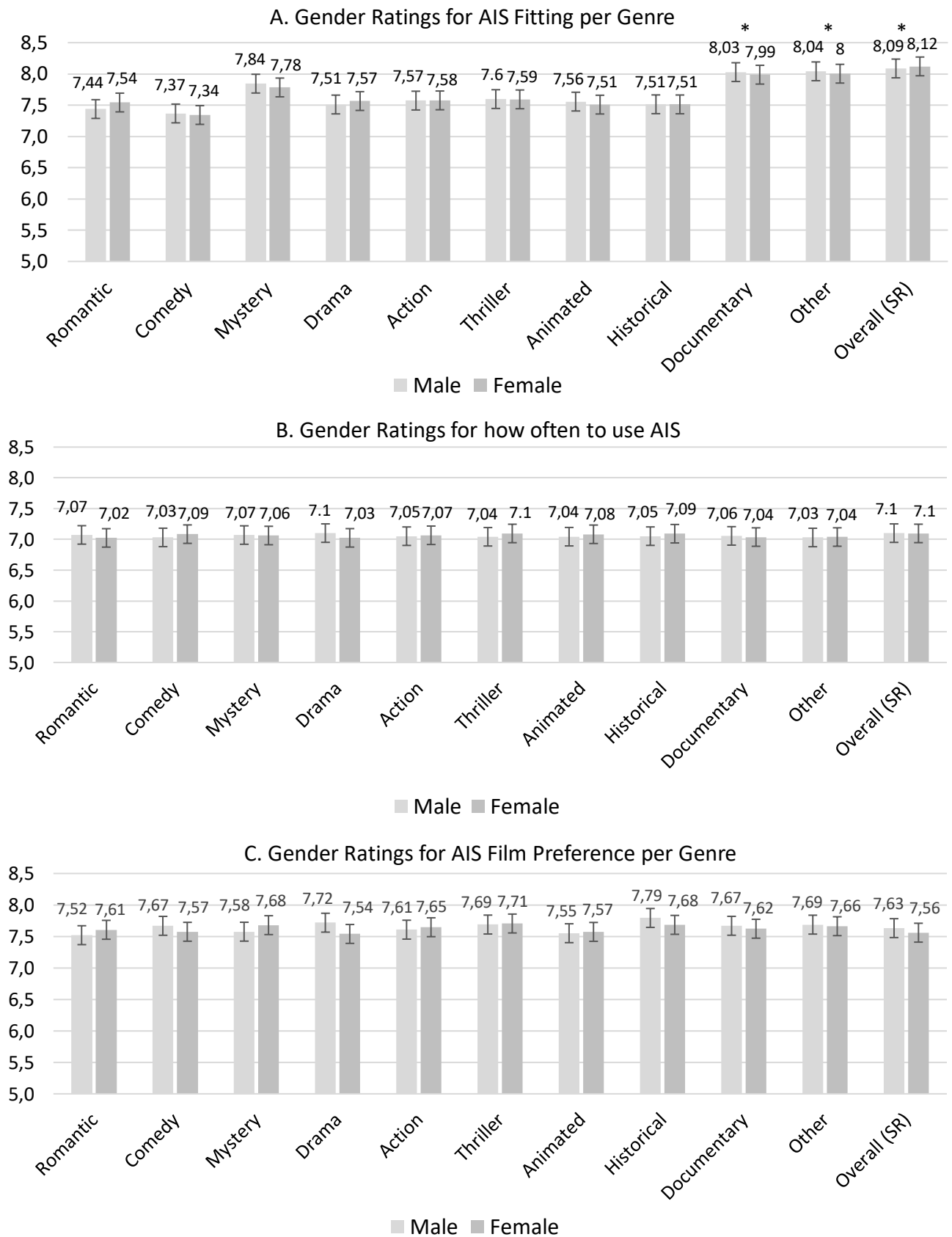


Figure 5: Between-gender and among-genres characteristics and comparisons for post-experimental questions (see Procedures; see Figure 2). Asterisks (*) over gender pairs in A. indicate significantly higher values for these items compared to other items at $p < .001$. No other effects were reported for these analyses (see also Figure 1). Bars indicate ± 2 standard errors from the mean.

4. Discussion

In this manuscript, we attempted to replicate the findings by Leveridge and colleagues (2024) and, critically, we explored the gender characteristics, and compared between genders for whether the presentation of a fantasy/thriller film elicited different physiological responses (SCR, HR) when the participants were presented with STS, AIS and no subtitles. We showed various key findings pertaining to this largely unexplored area in audiovisual translation including that AIS elicited the highest positively rated physiological correlates, that the no-subtitles condition elicited higher physiological changes than the STS condition, and that these findings were also reflected in self-report ratings for a multitude of engagement tasks, such as viewing quality and immersion in the clip context for both genders. Intriguingly, we showed that female participants experienced overall higher tonic and phasic physiological arousal as measured by SCR, and HR responses than male participants although this recorded arousal did not reflect in their self-report ratings.

Female participants experienced higher skin-conductance and heart-rate responses to the presented clips for AIS and STS and NS compared to male participants. This effect could be interpreted as an *expected* outcome given findings in previous research that have shown that female participants experience higher responses to fear than male participants in relevant studies (Timmers et al., 2003). Nevertheless, the current findings are novel in the sense that the effect is reported in a thriller-fantasy context involving diverse subtitling delivery methods and provides a response to topical reviews as to whether this effect could be replicated in Film Studies (Shields, 2003).

An interesting addition to topical research, as regards gender characteristics, is that although the differences in physiological responses between female (higher) and male (lower) participants were empirically indisputable from a statistical report perspective, the same did not apply for self-reports for valence and arousal. This is not the first time this effect is reported (see Leong et al., 2023) and highlights why the interplay of physiological responses, self-report ratings and gender comparisons has attracted so much attention in this area in recent years (see Blanco-Ruiz et al., 2020). It is entirely possible that the effect does not reflect a mere linear relationship and – although not suggested by previous research – that male and female audiences respond to self-report engagement tasks from a “physiological homeostasis” reference point. In simpler words, it is entirely possible that a male participant who experiences arousal in response to fearful stimuli self-assessed their arousal based on previous experiences and attributed a corresponding Likert-scale value to their experience. Conversely, a female participant might have experienced higher arousal to the same clip than a male participant, but their own reference point disturbed her “physiological homeostasis” equally when compared with a personal reference point to previous experiences, and, therefore, resulted in a proximate Likert-scale value that a male participant reported. This is an inherent problem with interval data, such as Likert scales (see van der Ploeg et al., 2017) in the sense that the reported outcomes can be subjective and the same Likert-scale value could indicate very different things between and among different individuals (see particularly, Jebb et al., 2021). Solutions to this problem can include parallel use of physiological and self-report Likert-scale responses – for which we opted in this manuscript, thus uncovering the seeming paradoxical effect at hand – and other less well-known and often understated methods, such as Rasch analyses (see Boone & Staver, 2020). For the purposes of this seminal outcome – beyond highlighting the importance of the currently applied, and further solutions – it should be noted that the importance that this effect confers for this area is that female participants can experience higher physiological responses than male participants in a thriller-fantasy film context, but that does not necessarily imply that they experience their physiological

responses as more extreme than male participants; they experience them from a self-assessment perspective that can have subjective value.

Furthermore to these, it should be noted that a very illustrative finding for the positive reception of AIS subtitles in the current research is that despite that pre-experimental ratings for how interested males and females were in watching different film genres yielded very significant differences (see Figure 1; see also Wuhr et al., 2017), these differences were not significant and provided Bayesian evidence for equivalence to the null for how interested male and female participants were to see different film genres with AIS (see Figure 5C.). This is a novel result and supports that AIS may play a very significant role in audiovisual translation in the years to come, but we must also note that given – particularly, the novelty of AIS delivery for female participants – it could be a curiosity effect, and of course as such it provided indicative but still inconclusive and pending for replication and further research outcomes for how often AIS should be used in films (see Figure 5B.).

5. Conclusions

The universal preference for the use of AIS in fiction genres is illustrated using short clips in the current manuscript. Whilst participants responded positively to the use of AIS, the process for making AIS still normally involves the use of video compositing software with a high start-up cost and level of expertise. The lack of accessible software and current workflows, whereby the production of subtitles is still largely an afterthought of the media-making process, are contributing factors to why AIS are not more widely available.

The present study demonstrates positive responses to the use of AIS in fiction. More studies on the reception of different subtitling deliveries could also go a long way towards helping to delineate ill-defined notions of engagement and immersion. Further research could also examine the use of AIS in non-fiction contexts, perhaps with a focus on comprehension, especially for material with a greater focus on informational content.

Whilst this study focusses on interlingual subtitles, the potential benefits (in terms of immersion, engagement or comprehension) of AIS in intralingual contexts largely remains to be explored. This is particularly important for members of the audience on the deafness spectrum, but also other demographics (particularly young people according to recent surveys carried out by Stagertext and YouGov) that are likely to enjoy watching material with subtitles. Evidence produced in the context of intralingual subtitles would also help determine whether the attraction for AIS is due to a perceived ability to mitigate for the foreignness of the material in interlingual contexts, or whether AIS simply dovetail nicely with audiovisual material. Answering such questions is critical towards making a case for a much broader use of AIS and for their integration into creative workflows.

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. <https://doi.org/10.1016/j.apmr.2018.07.066>

Alexithymia (2023, September 5). Alexithymia: Emotional blindness. Retrieved from <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. <https://doi.org/10.1521/bumc.2015.79.3.187>
- Ameri, S., Khoshsaligheh, M. & Khazaei Farid, A. (2018). The reception of Persian dubbing: a survey on preferences and perception of quality standards in Iran. *Perspectives*, 26(3), 435–451. <https://doi.org/10.1080/0907676X.2017.1359323>
- American Psychological Association. (2019). *Publication Manual of the American Psychological Association* (7th ed.). Arlington, VA: Author.
- Baker, D. H., Vilidaite, G., Lygo, F. A., Smith, A. K., Flack, T. R., Gouws, A. D., & Andrews, T. J. (2021). Power contours: Optimising sample size and precision in experimental psychology and human neuroscience. *Psychological Methods*, 26(3), 295–314. <https://doi.org/10.1037/met0000337>
- Balakrishnan, V. (2003). All about the Dirac delta function (?). *Resonance*, 8(8), 48–58. <https://doi.org/10.1007/BF02866759>
- 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>
- Bekmambetov, T. (Director). (2004). *Night Watch* [Motion picture]. Russia: Channel One Russia.
- Berryman, C., McAuley, J. H., & Moseley, L. G. (2012). Sphere 12 screening questionnaire. *Journal of Physiotherapy*, 58(4), 273–275. [https://doi.org/10.1016/S1836-9553\(12\)70133-9](https://doi.org/10.1016/S1836-9553(12)70133-9)
- Bianchin, M., & Angrilli, A. (2012). Gender differences in emotional responses: A psychophysiological study. *Physiology & Behavior*, 105(4), 925–932. <https://doi.org/10.1016/j.physbeh.2011.10.031>
- Birnbaum, D., Nosanchuk, T. & Croll, W. (1980). Children's stereotypes about sex differences in emotionality. *Sex Roles*, 6, 435–443. <https://doi.org/10.1007/BF00287363>
- 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>
- Blanco-Ruiz, M., Sainz-de-Baranda, C., Gutiérrez-Martín, L., Romero-Perales, E. & López-Ongil, C. (2020). Emotion elicitation under audiovisual stimuli reception: should artificial intelligence consider the gender perspective? *International Journal of Environmental Research and Public Health*, 17(22), 1–22. <https://doi.org/10.3390/ijerph17228534>
- Boone, W. J., & Staver, J. R. (2020). *Advances in Rasch analyses in the human sciences* (pp. 287–302). Springer.

- Bosseaux, C. (2023). *Experimental guidelines for subtitling emotional content in documentaries*. The Ethical Demands of Translating Gender-Based Violence: a Practice-Based Research Project Website. <http://ethicaltranslation.llc.ed.ac.uk/wp-content/uploads/2023/10/Experimental-guidelines-for-subtitling-emotional-content-in-documentaries-5.pdf>
- Bradley, M. M., Codispoti, M., Sabatinelli, D. & Lang, P. J. (2001). Emotion and motivation. II: Sex differences in picture processing. *Emotion*, 1(3), 300–319. <https://doi.org/10.1037//1528-3542.1.3.300>
- 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.
- Briton, N. & Hall, J. (1995). Beliefs about female and male nonverbal communication. *Sex Roles*, 32, 79–90. <https://doi.org/10.1007/BF01544758>
- 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>
- 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>
- 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. <https://doi.org/10.1177/2515245918823199>
- Cacioppo, J. T., Tassinary, L. G., & Berntson, G. (Eds.). (2007). *Handbook of psychophysiology*. Cambridge University Press.
- Chueh, T. H., Chen, T. B., Lu, H. H. S., Ju, S. S., Tao, T. H., & Shaw, J. H. (2012). Statistical prediction of emotional states by physiological signals with manova and machine learning. *International Journal of Pattern Recognition and Artificial Intelligence*, 26(04), 1250008. <https://doi.org/10.1142/S0218001412500085>
- Ciuk, D., Troy, A., & Jones, M. (2015). Measuring emotion: Self-reports vs. physiological indicators. 2015 Annual Meeting of the Midwest Political Science Association. <https://ssrn.com/abstract=2595359> or <http://dx.doi.org/10.2139/ssrn.2595359>
- 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>

- Dienes, Z. (2021). How to use and report Bayesian hypothesis tests. *Psychology of Consciousness: Theory, Research, and Practice*, 8(1), 9–26. <https://doi.org/10.1037/cns0000258>
- Di Giovanni, E. (2012). Italians and television: A comparative study of the reception of subtitling and voice-over. In S. Bruti & E. Di Giovanni (Eds.), *Audiovisual translation across Europe: An ever-changing landscape* (pp. 171–190). Peter Lang.
- Dong, C., & Wedel, M. (2014). BANOVA: an R package for hierarchical Bayesian ANOVA. *Journal of Statistical Software*, 10(2). <https://ssrn.com/abstract=2460286>
- Ekman, P. (1999). Basic emotions. In T. Dalgleish & M. Power (Eds.), *Handbook of Cognition and Emotion* (pp. 45–60). Wiley.
- Fabes R. A. & Martin C. L. (1991). Gender and age stereotypes of emotionality. *Personality and Social Psychology Bulletin*, 17(5), 532–540. <https://doi.org/10.1177/0146167291175008>
- 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. <https://doi.org/10.3758/BRM.41.4.1149>
- 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>
- Fischer, A. H. (1993). Sex differences in emotionality: Fact or stereotype? *Feminism Psychology*, 3(3), 303–318. <https://doi.org/10.1177/0959353593033002>
- Fox, W. (2016). Integrated titles: An improved viewing experience? In S. Hansen-Schirra & S. Grucza (Eds.), *Eyetracking and applied linguistics* (pp. 5–31). Berlin: 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*. PhD thesis. 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. <https://doi.org/10.1016/j.paid.2006.04.010>
- Graham, B. (2014). ‘John Wick’ Filmmakers on How to Direct Action, DGA, Keanu Reeves’ Advice, and More. *The Film Stage*. <https://thefilmstage.com/john-wick-filmmakers-on-how-to-direct-action-dga-keanu-reeves-advice-and-more/>

- Grossman, M. & Wood, W. (1993). Sex differences in intensity of emotional experience: A social role interpretation. *Journal of Personality and Social Psychology*, 65(5), 1010–1022. <https://psycnet.apa.org/doi/10.1037/0022-3514.65.5.1010>
- Gürkan, A. (2019). *Subtitling for the deaf and the hard-of-hearing: a reception study in the Turkish context*. PhD thesis. University College London.
- Hagemann, D., Naumann, E., Maier, S., Becker, G., Lürken, A., & Bartussek, D. (1999). The assessment of affective reactivity using films: Validity, reliability and sex differences. *Personality and Individual Differences*, 26(4), 627–639. [https://doi.org/10.1016/S0191-8869\(98\)00159-7](https://doi.org/10.1016/S0191-8869(98)00159-7)
- Handley, I. M., Brown, E. R., Moss-Racusin, C. A., & Smith, J. L. (2015). Quality of evidence revealing subtle gender biases in science is in the eye of the beholder. *Proceedings of the National Academy of Sciences*, 112(43), 13201–13206. <https://doi.org/10.1073/pnas.1510649112>
- 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.
- Jebb, A. T., Ng, V., & Tay, L. (2021). A review of key Likert scale development advances: 1995–2019. *Frontiers in psychology*, 12, 637547. <https://doi.org/10.3389/fpsyg.2021.637547>
- Kelly, M. M., Tyrka, A. R., Anderson, G. M., Price, L. H. & Carpenter, L. L. (2008). Sex differences in emotional and physiological responses to the Trier social stress test. *Journal of Behavior Therapy and Experimental Psychiatry*, 39(1), 87–9. <https://doi.org/10.1016/j.jbtep.2007.02.003>
- Keselman, H. J., Huberty, C. J., Lix, L. M., Olejnik, S., Cribbie, R. A., Donahue, B., ... & Levin, J. R. (1998). Statistical practices of educational researchers: An analysis of their ANOVA, MANOVA, and ANCOVA analyses. *Review of educational research*, 68(3), 350–386. <https://doi.org/10.3102/003465430680033>
- 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. <https://doi.org/10.1177/1073191119837613>
- 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>

- Kring, A. M. & Gordon, A. H. (1998). Sex differences in emotion: Expression, experience, and physiology. *Journal of Social and Personality Psychology*, 74(3), 686–703. <https://psycnet.apa.org/doi/10.1037/0022-3514.74.3.686>
- 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). Amsterdam and Philadelphia: John Benjamins Publishing Company.
- Labouvie-Vief, G., Lumley, M. A., Jain, E. & Heinze, H. (2003). Age and gender differences in cardiac reactivity and subjective emotion responses to emotional autobiographical memories. *Emotion*, 3(2), 115–126. <https://doi.org/10.1037/1528-3542.3.2.115>
- Leong, M. Q., Yu, Z., Tsikandilakis, M., & Tong, E. M. W. (2023). “See no evil. Feel no evil?”: Exploring emotional responses to masked moral violations in religious and nonreligious Singaporean participants. *Evolutionary Behavioral Sciences*. Advance online publication. <https://doi.org/10.1037/ebs0000333>
- Levenson, R.W., Carstensen, L.L., Friesen, W.V., & Ekman, P. (1991). Emotion, physiology, and expression in old age. *Psychology and Aging*, 6, 28–35. <https://doi.org/10.1037/0882-7974.6.1.28>
- Levenson, R.W., Ekman, P., & Friesen, W.V. (1990). Voluntary facial action generates emotion specific autonomic nervous system activity. *Psychophysiology*, 27, 363–384. <https://doi.org/10.1111/j.1469-8986.1990.tb02330.x>
- Leveridge, F., Mevel, 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>
- Lithari, C., Frantzidis, C.A., Papadelis, C., Vikas, A. B. & Klados, M. A. (2010). Are Females More responsive to emotional stimuli? A neurophysiological study across arousal and valence dimensions. *Brain Topography*, 23, 27–40 <https://doi.org/10.1007/s10548-009-0130-5>
- Maffei, A., Vencato, V. & Angrilli, A. (2015). Sex differences in emotional evaluation of film clips: interaction with five high arousal emotional categories. *PLoS ONE*, 10(12), 1–13. <https://doi.org/10.1371/journal.pone.0145562>
- Mayer, L. S., & McHugh, P. R. (2016). Sexuality and gender: Findings from the biological, psychological, and social sciences. *The New Atlantis*, 50, 10–143.
- McClarty, R. (2012). Towards a multidisciplinary approach in creative subtitling. *MonTI*, 4(4), 113–153. <https://doi.org/10.6035/MonTI.2012.4.6>

- Orrego-Carmona, D. (2016). A reception study on non-professional subtitling: Do audiences notice any difference? *Across Languages and Cultures*, 17, 163–181. <https://doi.org/10.1556/084.2016.17.2.2>
- Plant, E. A., Hyde, J. S., Keltner, D. & Devine, P. G. (2000). The gender stereotyping of emotions. *Psychology of Women Quarterly*, 24(1), 81–92. <https://doi.org/10.1111/j.1471-6402.2000.tb01024.x>
- Plevkova, J., Brozmanova, M., Harsanyiova, J., Sterusky, M., Honetschlager, J., & Buday, T. (2020). Various aspects of sex and gender bias in biomedical research. *Physiological research*, 69 (Suppl 3), S367. <https://doi.org/10.33549/physiolres.934593>
- 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. <https://doi.org/10.1016/j.biopsycho.2017.08.051>
- 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>
- Ramos, M. (2015). The emotional experience of films: does Audio Description make a difference? *The Translator*, 21(1), 68–94. <https://doi.org/10.1080/13556509.2014.994853>
- Rawsthorn, A. (2007). The Director Timur Bekmambetov Turns Film Subtitling into an Art. *The New York Times*. https://www.nytimes.com/2007/05/25/style/25iht-design28.1.5866427.html?_r=0
- Rohrmann, S., Hopp, H., & Quirin, M. (2008). Gender differences in psychophysiological responses to disgust. *Journal of Psychophysiology*, 22(2), 65–75. <https://doi.org/10.1027/0269-8803.22.2.65>
- Romero-Fresco, P. (2019) *Accessible filmmaking: Integrating translation and accessibility into the filmmaking process*. Routledge.
- Romero-Fresco, P., & Fryer, L. (2018). *Accessible filmmaking guide*. Archer's Mark.
- Schäfer, T., & Schwarz, M. A. (2019). The meaningfulness of effect sizes in psychological research: Differences between sub-disciplines and the impact of potential biases. *Frontiers in psychology*, 10, 442717. <https://doi.org/10.3389/fpsyg.2019.00813>
- Shields, S. A. (2003). *Speaking from the heart: Gender and the social meaning of emotion*. Cambridge University Press.
- Sun, T., Gaut, A., Tang, S., Huang, Y., ElSherief, M., Zhao, J., ... & Wang, W. Y. (2019). Mitigating gender bias in natural language processing: Literature review. *arXiv preprint arXiv:1906.08976*. <https://doi.org/10.48550/arXiv.1906.08976>

- 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>
- Thunberg, M. & Dimberg, U. (2000). Gender differences in facial reactions to fear-relevant stimuli. *Journal of Nonverbal Behavior*, 24, 45–51. <https://doi.org/10.1023/A:1006662822224>
- Timmers, M., Fischer, A. & Manstead, A. (2003). Ability versus vulnerability: Beliefs about men's and women's emotional behaviour. *Cognition & Emotion*, 17(1), 41–63. <https://doi.org/10.1080/026999303022277>
- 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. <https://doi.org/10.1016/j.concog.2017.10.013>
- 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. <https://doi.org/10.1016/j.concog.2019.102771>
- Tsikandilakis, M., Bali, P., Derrfuss, J., & Chapman, P. (2020). “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. <https://doi.org/10.1080/02699931.2019.1646710>
- Tsikandilakis, M., Leong, M. Q., Yu, Z., Paterakis, G., Bali, P., Derrfuss, J., ... & Mitchell, P. (2021). “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. <https://doi.org/10.1007/s00426-020-01461-7>
- Tsikandilakis, M., Bali, P., Karlis, A., Mével, P. A., Madan, C., Derrfuss, J., & Milbank, A. (2023). Unbiased individual unconsciousness: Rationale, replication and developing applications. *Current Research in Behavioral Sciences*, 100109. <https://doi.org/10.1016/j.crbeha.2023.100109>
- 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>.
- von Flotow, L. & Josephy-Hernández, D. E. (2018) Gender in audiovisual translation: Advocating for gender awareness. In L. Pérez González (Ed.), *The Routledge Handbook of Audiovisual Translation* (pp. 296–311). Routledge.

Vrana, S. R. & Rollock, D. (2002). The role of ethnicity, gender, emotional content, and contextual differences in physiological, expressive, and self-reported emotional responses to imagery. *Cognition & Emotion*, 16(1), 165–192. <https://doi.org/10.1080/02699930143000185>

Wühr, P., Lange, B. P., & Schwarz, S. (2017). Tears or fears? Comparing gender stereotypes about movie preferences to actual preferences. *Frontiers in psychology*, 8, 249605. <https://doi.org/10.3389/fpsyg.2017.00428>