

Audio subtitling and subtitling: a comparison of their emotional effect on blind / partially sighted and sighted users

Gonzalo Iturregui-Gallardo

Universitat Autònoma de Barcelona
Espanya

Olga Soler-Vilageliu

Universitat Autònoma de Barcelona
Espanya

ONOMÁZEIN | Special Issue VIII – Emotions in Translation and Interpreting: 61-82

DOI: 10.7764/onomazein.ne8.07

ISSN: 0718-5758



Gonzalo Iturregui-Gallardo: Universitat Autònoma de Barcelona, Department of Translation, Interpretation and East Asian Studies. Transmedia Catalonia Group. ORCID: 0000-0003-3664-0045

| E-mail: gonzalo.iturregui@uab.cat

Olga Soler-Vilageliu: Universitat Autònoma de Barcelona, Department of Psychology, Department of Basic, Evolutionary and Educational Psychology. Transmedia Catalonia Group. ORCID: 0000-0001-9219-1913

| E-mail: olga.soler@uab.cat

Received: July 8, 2020

Accepted: October 2, 2020

Abstract

Audio subtitling (AST) is a media accessibility service that allows for people who are blind, partially sighted or with any reading disability to access written subtitles in their aural form. Despite the existing literature on other media accessibility services such as audio description, the way written subtitles compare to orally delivered subtitles has not been researched. In this study, a group of 42 blind and partially sighted participants and a group of 42 sighted participants watched the same three video clips. Two of them pictured two emotions (sadness and fear) and the third was emotionally neutral. The clips were prepared with subtitles or audio subtitles, according to the target participants. The emotional effect of the clips was measured in two ways: with self-reports, by completing after each clip the SAM questionnaire (Bradley & Lang, 1994); and with psychophysiological measures: electrodermal activity (EDA) and heart rate (HR), that were recorded while participants watched each clip. The analyses of the data obtained in both experiments indicate that self-report measures revealed similar experiences for both blind and sighted participants, differentiating between valence (negative for fear and sadness, but not for neutral) and arousal (deemed higher for fear than for sadness and neutral). Data from EDA and HR measures are less conclusive.

Keywords: media accessibility; psychophysiology; audio subtitling; emotional arousal; SAM questionnaire.

1. Introduction

This article presents a study on the reception of audiovisual contents. Most particularly, the research conducted is found within the fields of audiovisual translation and media accessibility. We used measures from the field of psychology to put in contrast the reception of audiovisual contents in their standard form (subtitles) and their accessible form (audio subtitles).

There is a consistent body of literature on the main transfer modes, which are dubbing, voice-over or subtitling. Proof of it are recent studies by Perego, Del Missier, Porta, & Mosconi (2010) and Kruger, Soto-Sanfiel, Doherty, & Ibrahim (2016) on subtitling; by Matala, Perego, & Bottiroli (2017) and Perego, Del Missier, & Bottiroli (2014) on the comparison of dubbing and subtitling reception; and by Sepielak (2016) on subtitling and voice-over. Audio description has also been studied in depth, mainly in cinema (Walczak & Fryer, 2018; Walczak, 2017; Wilken & Kruger, 2016), but in theatre (Udo, Acevedo & Fels, 2010) and opera (Eardley-Weaver, 2014) as well. Conversely, only few studies have focused on audiosubtitling, to our knowledge Braun & Orero (2010), Orero (2007), Benecke (2012), Remael (2012), Orero (2011), Reviere & Remael (2015), Thrane (2013). Our study intends to contribute to the understanding of this accessibility service.

2. Audio subtitling and subtitling

Subtitling is a transfer mode that allows for access to those people who have difficulty understanding the original language of a content or people who are deaf or deaf of hearing (Gambier, 2006). This mode is based on the incorporation of written text on screen in the language known to the audience where an original version of the audiovisual content is shown. Subtitles coincide approximately with the dialogues delivered by the actors in the content (Chaume, 2010). It is one of the main transfer modes for the translation of audiovisual contents, together with dubbing (Díaz Cintas, 1999). In the field of translation studies, many scholars have produced research on the topic (see, for example, Bartoll, 2006; Chaume, 2003; Díaz Cintas, 2001; Díaz Cintas & Remael, 2007).

However, as text on screen, subtitling can pose a problem for people who are blind or who are partially sighted. Audio subtitling stands as an accessibility service, related to that of audio description (AD), which allows for an aural version of text on screen. The audio subtitling service is defined by the general concept of “spoken or aural rendering” of written subtitles (Fryer, 2013; Miesenberger, Klaus, & Zagler, 2002; Nielsen & Bothe, 2008; Orero, 2011; Mączyńska, 2011).

Accordingly, the original language is translated into another language in the written subtitles and then these are delivered orally. The way audio subtitles are delivered provides different information about the message or the multilingual reality of the audiovisual content they are

applied to. As suggested by Iturregui-Gallardo (2018), by applying some variations in aspects such as the volume, the acting, or the information provided by the audio describer, audio subtitles can adopt different forms.

Audio subtitling is also deemed to help people with needs other than the blind and partially sighted. Some examples are consumers of audiovisual materials, including elderly people or people with language and reading impairments—e.g., aphasia, dyslexia (Movietalk, 2014)—or cognitive impairment such as intellectual disability or decreased concentration (Braun & Otero, 2010; Ljunglöf, Derbring, & Olsson, 2012). This accessibility service could also help people who do not have any kind of visual impairment or any type of sensory or intellectual disability, such as children who have not yet entirely developed their reading skills and learners of a second language (Mihkla Hein, Kiissel, Rapp, Sirts, & Valdna, 2014).

3. Emotional arousal and psychophysiology in audiovisual experiences

The study of emotions has been around for more than 100 years in the field of research, as noted by Kleinginna & Kleinginna (1981). These authors collected more than 90 definitions and statements. As a result, they proposed the following definition, that intends to bring together the different components of emotion:

Emotion is a complex set of interactions among subjective and objective factors, mediated by neural hormonal systems, which can (a) give rise to affective experiences such as feelings of arousal, pleasure/displeasure; (b) generate cognitive processes such as emotionally relevant perceptual effects, appraisals, labelling processes; (c) activate wide spread physiological adjustments to the arousing conditions; and (d) lead to behaviour that is often, but not always, expressive, goal-directed and adaptive (Kleinginna & Kleinginna, 1981: 355).

This wide definition of Kleinginna & Kleinginna (1981) actually points out two different types of measures for emotion: those related to the affective experiences and cognitive processes, that can be called *subjective* measures of emotion, and those related to the measurement of psychophysiological responses as *objective* measures, which, according to Kreibig (2010), seem to be a flourishing field in many scientific disciplines. A diversity of questionnaires aimed at the cognitive evaluation of the emotional state have been proposed by different scholars. Among others, and as examples, see the Adjective Checklist (MACL) (Nowlis, 1965), the Positive and Negative Affect Schedule (PANAS-X) (Watson & Clark, 1999), or the Self-Assessment Manikin (Bradley & Lang, 1994), which was used in the present study.

On the other hand, changes in the physiology of the organism derived from psychological processes have proved to relate to what people feel and experience. When recorded, these changes can provide more objective data on emotions. Several arousal indicators can be used to measure the organism's reaction: brainwaves, muscle contraction, electrodermal activity (EDA) or heart rate (HR). EDA is related to the amount of sweat located in the skin,

which will make the electric skin conductance higher, the more sweat is produced. On the other hand, HR registers in beats per second, which may vary in relation to the stimuli to which a person is exposed.

The combination of EDA and HR is a usual one. Kettunen et al. (1998) stated that there is synchronization between EDA and HR, and in their study both measures correlated with self-report measures. These two measures were combined in some studies with different audiovisual stimuli, such as immersive environments (Drachen et al., 2010; Egan et al., 2016), cinema (Rooney et al., 2012; Dillon, 2006), audio described and audio subtitled contents (Fryer, 2013; Iturregui-Gallardo et al., 2018), videogames (O'Hagan, 2010, 2016), video frame segmentation (Wilson & Sasse, 2006), or interactive technologies (Mandryk, Inkpen, & Calvert, 2006).

In the study of the emotional effects of audiovisual stimuli, emotion relates to the experience of entertainment, described as pleasurable experience. Entertainment is linked here to the psychological processes of flow (Csikszentmihalyi, 1997), and most particularly media flow (Sherry, 2004), which can be summarised as a mental state characterised by absorption and concentration in an activity, in this case media consumption, and presence {Formatting Citation}. From a theoretical point of view, the relation between the conception of media as a means of entertainment and the psychological processes of flow and presence has been also developed in previous works (Bartsch, Mangold, Viehoff, & Vorderer, 2006; Dillon, 2006; Dillon et al., 2001; Klimmt & Vorderer, 2003; Mikropoulos, Tzimas, & Dimos, 2004; Reeves, Lang, Kim, & Tatar, 2009; Simons, Detenber, & Roedema, 1999).

Particularly in the field of Media Accessibility (MA), the impact of AD on blind and partially sighted audiences—primary receivers and users of AD—has only been studied by means of physiological markers, to the best of our knowledge, in two studies. First, HR and EDA (or galvanic skin response) were used by Fryer (2013) in combination with questionnaires on presence to test the immersion and the Elicited Emotion Scale developed by Gross and Levenson (1995) of audio described films with no AD, synthetic AD, and human AD. Later, HR was used by Ramos (2015) in combination with the PANAS-X questionnaire (Watson & Clark, 1999) to test the emotional activation experience by participants when exposed to audio described or non-audio described emotional clips. In Fryer (2013), 19 blind and partially sighted participants (M=48.53) took part in the experiment; each of the participants was exposed to six clips, 2 for each condition. The procedure combined the presentation of stimuli with 30-second relaxation periods. In Ramos (2015), a total of 70 participants took part in the experiment: 30 were blind and partially sighted (~34 years old) and 40 were sighted (~34 years old), of which 10 were blindfolded. In this experiment, the group of sighted participants was also exposed to audio described clips. The relaxation period between the stimuli was of 60 seconds. The results of both studies presented great differences between self-report questionnaires and HR. Moreover, HR presented some alterations related to the stress caused by the experimental context.

The experiments here reported are part of a study developed within the NEA project (New Approaches to Accessibility, 2015-2018, [FFI2015-64038-P]) at the Autonomous University of Barcelona. This study compared the reception of the same audiovisual content by two different target audiences: a blind and partially sighted users and sighted users. The audiovisual stimuli used were the same, adapted for each of the groups in their subtitled and audio subtitled versions. The main goal of the study was to assess the differences on emotional reactions between the groups. In addition, the experiment aimed at exploring new measuring methods with the main purpose of attaining more objectivity, based on previous studies within media accessibility such as the above mentioned of Fryer (2013) and Ramos Caro (2015, 2016). Our study used two key measures in psychophysiology, EDA and HR, combined with a self-report questionnaire.

4. Experiment 1—blind and partially sighted participants

4.1. Method

This experiment combined self-report and psychophysiological measures to test the emotional arousal experienced by the blind and partially sighted participants when exposed to audio subtitled clips.

4.1.1. Participants

The experiment was performed with blind (13) and partially sighted (29) participants (n=42). 17 of them were women. The age mean was of 38 years old. Most of the participants were contacted through an association of people with sight loss, B1 B2 B3 (<https://www.b1b2b3.org/ca/>). Each participant was asked to fill a consent form. The experiment and the forms to be distributed to the participants had the approval by the ethical committee of the Autonomous University of Barcelona.

4.1.2. Material

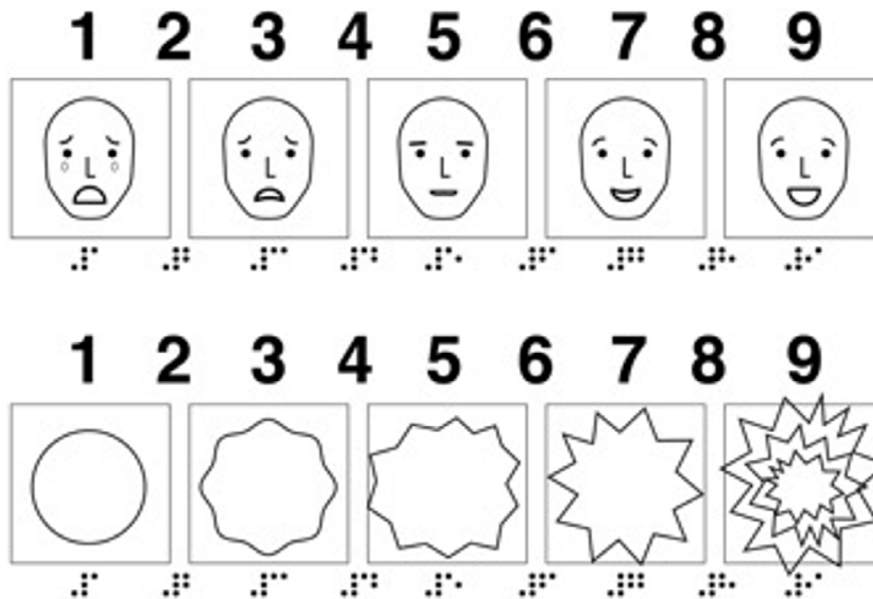
4.1.2.1. Instruments

Self-report. The T-SAM questionnaire consists of a tactile, simplified, and augmented version of the SAM questionnaire. The original questionnaire, designed in the 90s by Bradley & Lang (1994), is based on three fundamental dimensions of emotion: valence (the positive or negative emotional value attributed to the situation or stimulus that triggers the emotion), arousal (from not activating at all to extremely activating), and dominance (from no control to complete control of the emotion). The questionnaire is constituted by three subscales where the different levels (from minimum to maximum) of emotional valence, arousal and dominance, are represented by an illustration accompanied by a numerical 9-point scale. Due to its brevity, simplicity, and transcultural character, the SAM represents one of the gold standards

in emotional evaluation (Bradley, Codispoti, Cuthbert, & Lang, 2001a; Bradley, Codispoti, Sabinelli, & Lang, 2001b; McManis, Bradley, Berg, Cuthbert, & Lang, 2001).

FIGURE 1

Image of the Tactile-Self Assessment Manikin questionnaire



First of all, for the adapted version of the SAM, the third dimension of dominance was removed, as suggested in more recent works (Lang, Bradley, & Cuthbert, 2008; Montefinese, Ambrosini, Fairfield, & Mammarella, 2014). The experiments combined the questionnaire with psychophysiological measures that find correspondence in the two first scales, in particular the second one. It was then adapted to a simplified, augmented, and tactile version that aims to suit participants with different kinds of requirements. Two different proposals of design were vectorized and then printed with UV gloss on a non-absorbing surface (PVC), and they were presented to blind and partially sighted participants in a focus group in order to develop a final version with the suggestions and corrections that arose in the session (Iturregui-Gallardo & Méndez-Ulrich, 2019). Figure 1 above shows the final design of the T-SAM.

Psychophysiology. Two different psychophysiological markers were recorded in this experiment: EDA and HR. The CAPTIV L7000 system was used to record both measurements. The set is composed by a central device, which is connected to a computer and receives wireless data from the sensors. The first sensor, for EDA, splits in two sensors that are attached under the second phalange of the pointing and middle fingers; the second, for HR, is a belt that is placed below the chest, at the end of the sternum. Data were recorded at a frequency of 32 Hz. The recording software allows for the recording and visualisation of data. This software was linked to the Tobii Studio software, which worked as experiment controller.

4.1.2.2. Stimuli

Scenes from the Polish TV series *Wojenne Dziewczyny* [War Girls]¹ (2017), broadcast by the Polish national channel *Telewizja Polska 1* (TVP1), were used. Polish was selected since the language had to be unknown to the audience in order to reproduce a multilingual content. A copyright agreement was reached with the broadcast company. Scenes should have a duration longer than 2 minutes and be self-contained in terms of meaning, in order to involve the audience. Initially, eight scenes were chosen. Negative emotions were selected since they are easier to induce than positive ones and create stronger reactions in the organism (Taylor, 1991; Uhrig et al., 2016; Westermann, Spies, Stahl, & Hesse, 1996) and are the most used in emotion research (Kreibig, 2010). The inclusion of a neutral clip had been suggested in previous studies on emotion with the use of psychophysiological measures (Kreibig, Wilhelm, Roth, & Gross, 2007). A professional translator subtitled the clips in Spanish. An online validation, using the SAM questionnaire with more than 120 sighted respondents, served to select the three final emotional clips: fear, sadness, and neutral. Each of the clips was treated in a professional studio by professional dubbing actors, who were told to register with a neutral reading-aloud style. Audio description was commanded to a professional audio describer. The AD and AST track was recorded in a dubbing studio and delivered by voice talents.

4.1.3. Procedure

The participants were asked to fill a consent form read orally to them. Then some preliminary demographic questions were asked. The sensors were placed, the set was tested and the procedure was initiated. All the instructions were presented as slides on the screen and delivered aurally.

The presentation started with the description of the tasks and the introduction of the TV series and the plot. Before the presentation of the first clip, the participant was induced into a relaxation period of 8 minutes with relaxing music. EDA and HR were recorded during this period and as a baseline we took the mean values of the last 3 minutes. After the sound of a bell, a short description of the scene was provided. Clips were randomly presented. After each of the clips, the participants responded orally to the T-SAM questionnaire and the researcher wrote down the answers. After the first and second clip, a relaxation period of 3 minutes without music was induced. The procedure lasted about 50 minutes, including the presentation of the pre-test and a preference test at the very end to gather the subjective opinion of the users on audio subtitling features such as voice or prosody.

1 <https://vod.tvp.pl/website/wojenne-dziewczyny,28767487>.

4.2. Results

The first section presents the results for the self-report questionnaire T-SAM, the second focuses on the results obtained from psychophysiological measurements.

4.2.1. T-SAM Questionnaire

Valence. The median values obtained for the valence dimension were 3 for sadness, 3 for fear and 5.5 for neutral (see Figure 2). Since T-SAM's results are an ordinal value (not continuous), it is advisable to perform non-parametric tests to check the differences between the ratings for each emotion-eliciting clip. A Related-Samples Friedman's Two-Way Analysis of Variance by Ranks indicated significant differences (Friedman's $Q_{(2)} = 24,7$; $p = ,000$, $N = 42$). Pairwise comparisons indicate that the ratings for sadness and fear do not differ significantly (Friedman's $Q_{(1)} = 1,581$; $p = ,209$), whereas both ratings are different for the ratings of neutral: Sadness/Neutral: Friedman's $Q_{(1)} = 25,0$; $p = ,000$; Fear/Neutral: Friedman's $Q_{(1)} = 9,529$; $p = ,002$.

Arousal. The results for arousal were 5 for sadness, 7 for fear, and 4 for neutral. The Related-Samples Friedman's Two-Way Analysis of Variance by Ranks confirmed that the observed differences are significant (Friedman's $Q_{(2)} = 41,367$; $p = ,000$). Pairwise comparisons indicated that the difference between ratings for sadness and neutral are not statistically significant (Friedman's $Q_{(1)} = 3,789$; $p = ,052$), but that fear and neutral ratings did yield a significant difference (Friedman's $Q_{(1)} = 21,900$; $p = ,000$) as well as the ratings for fear and sadness (Friedman's $Q_{(1)} = 23,684$; $p = ,000$).

4.2.2. Psychophysiological measures

EDA. Due to the technical malfunctioning of sensors obtaining EDA measures, the data of 10 participants could not be included. The average for the values obtained was calculated. Later, the difference between the average minus the average of the baseline was calculated. A General Linear Model of Repeated measures comparing the mean values obtained for each clip did not reach statistical significance ($F_{(2,64)} = .212$; $p = .809$, Partial Eta Squared = .007).

Heart Rate. Due to the technical malfunctioning of sensor measuring HR, the data of 9 participants could not be included. In order to compare the effects of emotion on HR, the data were prepared in the same way as for EDA. The General Linear Model of Repeated measures analyses provided significant differences for HR in terms of emotion ($F_{(2,64)} = 3.320$, $p = .042$; Partial Eta Squared = .94). Pairwise analyses indicated that the values for sadness and neutral are different ($t_{(32)} = -2.40$; $p = .021$), and differences between sadness and fear almost reached significance ($t_{(32)} = -1,973$, $p = .057$).

4.2.3. Data correlation

In order to test the relationship between the results obtained by means of the self-report instrument and psychophysiological measures a Spearman Rank Correlation was run. The

relations are made between the physiological measures and the ratings of arousal for each type of stimuli presented to the users. No correlation was found between these two sets of data. For sadness, the relationship of EDA and arousal showed very low correlation ($r_s(31) = .14, p = .42$), and the same happens between HR and arousal ($r_s(31) = .01, p = .91$). For fear, the relationship of EDA and arousal showed again very low correlation ($r_s(31) = -.10, p = .57$), and the same is found between HR and arousal ($r_s(31) = -.32, p = .06$). For neutral, the relationship of EDA and arousal showed no correlation as well ($r_s(31) = .10, p = .57$), and it is the same between HR and arousal ($r_s(31) = .03, p = .85$).

4.3. Discussion

Significant differences among scenes were found for valence: the ratings of sadness (3) and fear (3) were different from those obtained for neutral (5.5). These results were expected and correspond to the description of emotions in which sadness and fear are considered negative emotions. Similarly, for arousal some differences were significant. The ratings of the clips for sadness (5) and neutral (4) do not present significant differences; however, they differ from the ratings of fear (7). Fear is considered an emotion causing high arousal, followed by sadness, which is also negative in valence but results in less arousal. Therefore, emotions were appropriately induced by the clips.

Regarding physiological measures, no significant differences were found for EDA, as in previous experiences (Fryer, 2013). For the results obtained from HR, some differences were observed. The values obtained with sadness were the lowest, statistically different from neutral and very close to be significantly different from fear. According to these results, HR slowed when the sad clip was shown. In previous experiments with AD and blind and partially sighted participants, Fryer (2013) did not report any significant result when studying HR, and Ramos (2015) reported that arousal in HR did not correspond with emotional arousal. The results here obtained showed HR deceleration for all emotions, which can be linked to an increase in concentration (Graham, 1992; Petrie Thomas, Whitfield, Oberlander, Synnes, & Grunau, 2012; Tremayne & Barry, 2001).

To sum up, no correlations were found between the numerical values assigned to the arousal scale in the SAM and the physiological measures that indicate actual arousal of the body. This finding indicates that participants cognitively realised the emotion they had to interpret but did not get sufficiently engaged as to have a physical reaction to it.

5. Experiment 2—sighted participants

This experiment was performed with the same materials and the same measures as experiment 1. As the participants in the experiment were sighted, a different treatment had to be applied to the clips, that were presented with current subtitles.

5.1. Method

The same combination of self-report and psychophysiological measures was used to test the emotional arousal experienced by sighted participants when exposed to subtitled clips.

5.1.1. Participants

42 sighted participants took part in Experiment 2. 27 were women. The mean age was 30 years old. The group of sighted participants were contacted through social media and other kind of advertising, such as mailing lists and other kinds of networking tools.

5.1.2. Material

5.1.2.1. Instruments

Self-report. As it has been stated before, the same self-report questionnaire, the T-SAM, was used in this Experiment 2. The T-SAM is an adaptation of the SAM questionnaire (Bradley & Lang, 1994), which was not only performed to make it tactile but also simplified and augmented.

Psychophysiology. Again, the same psychophysiological measurements were applied in this stage of the study. EDA and HR were measured by means of the sensors and the device Captiv L7000.

5.1.2.2. Stimuli

The same clips were subtitled in Spanish by a professional translator.

5.1.3. Procedure

The participants were asked to fill a consent form. Then some preliminary demographic questions were asked. The sensors were placed, the set tested, and the presentation of procedure began. All the instructions were presented as slides on the screen and delivered aurally. The presentation of the stimuli was combined with relaxation periods that were aimed at the registering of a baseline for each participant and the administration of the self-report questionnaire. With sighted participants, the procedure lasted also about 40 minutes, including the pre-questionnaire. They did not complete a preference test, like participants in Experiment 1 did, since they watched the clips in their original form with subtitles.

5.2. Results

5.2.1. T-SAM questionnaire

Valence. The results (in medians) obtained were 4 for sadness, 3 for fear, and 6.5 for neutral (see Figure 2). As in Experiment 1, non-parametric tests were performed to analyse the differ-

ences among the ratings for each emotion-eliciting clip. The Related-Samples Friedman's Two-Way Analysis of Variance by Ranks indicated that the clips were rated differently (Friedman's $Q_{(2)} = 56.08$; $p = .000$, $N = 42$). Pairwise comparisons yielded that the ratings of sadness and fear did not show significant differences (Friedman's $Q_{(1)} = 1$; $p = .317$). The ratings for both sadness and fear were different from neutral (Sadness/Neutral: Friedman's $Q_{(1)} = 41.0$; $p = .000$; Fear/Neutral: Friedman's $Q_{(1)} = 32.4$; $p = .000$).

Arousal. The results for arousal were 4 for sadness, 7 for fear, and 3 for neutral. A Friedman's Two-Way Analysis of Variance by Ranks confirmed the significance (Friedman's $Q_{(1)} = 50.302$; $p = .000$). Pairwise comparisons indicated that there are significant differences between sadness and neutral (Friedman's $Q_{(1)} = 6.429$; $p = .011$). Values for sadness and fear were also different (Friedman's $Q_{(1)} = 24.381$; $p = .000$). Finally, differences between fear and a neutral were significant too (Friedman's $Q_{(1)} = 36.100$; $p = .000$).

5.2.2. Psychophysiological measures

EDA. As happened in Experiment 1, some of the EDA sensors did not function properly and the recording of 6 participants could not be included. The results are shown in Figure 4 below. As with Experiment 1 data, a General Linear model of Repeated Measures was used to analyse the possible differences. The overall comparison did not show differences ($F_{(2,70)} = 2.110$, $p = .148$, Partial Eta Squared = .057), although pairwise comparisons yielded significant differences between sadness and fear EDA's ($t_{(35)} = -2.717$, $p = .010$).

Heart Rate. Again, the sensor for the recording of heart rate did not work well and the recordings of 11 participants had to be excluded from the analysis. The results obtained are shown in Figure 5 below. The General Linear Model of Repeated Measures applied to the values of the recording of heart rate did not show any significant difference ($F_{(2,60)} = .281$, $p = .756$, Partial Eta2 = .009).

5.3. Discussion

The results obtained from the ratings of the T-SAM questionnaire yielded significant differences. For valence, differences were observed between fear and sadness, and neutral. According to the description of emotions, both fear and sadness are considered negative emotions, thus implying that participants were induced the targeted emotions. As for arousal, results also yielded significance. Neutral caused less arousal (3), followed by sadness (4) and fear (7). Results are coherent with the description of emotions, which present fear as a very arousing emotion in relation to sadness.

For EDA, significant differences were only found between sadness and fear. According to the results, fear causes more activation and is remarkable higher than in sadness clip. On the other hand, values obtained HR did not yield any significant differences. As in Experiment

1, correlations were calculated with the Spearman Rank Correlation Coefficient applied to nonparametric measures. The correlations were made between each physiological measure and arousal ratings. All data seem to show very little correlation. For sadness, the correlation of EDA and arousal showed very low results ($r_s(34) = -.24, p = .15$), and the same goes between HR and arousal ($r_s(28) = .11, p = .56$). For fear, the correlation between EDA and arousal showed very low values ($r_s(34) = -.19, p = .26$), and the same is true between HR and arousal ($r_s(28) = .09, p = .60$). For neutral, the correlation between EDA and arousal also showed low values ($r_s(34) = .02, p = .87$) as well as between HR and arousal ($r_s(28) = -.007, p = .97$). In general, correlation has not shown any statistic similitudes between psychophysiological measures and the results obtained in the self-report instrument for any of the emotions in any of the groups.

6. Comparison of results Experiment 1 and Experiment 2

First, this section presents the comparison of the results obtained in the T-SAM questionnaire, and later we compare psychophysiological measures, EDA, and HR.

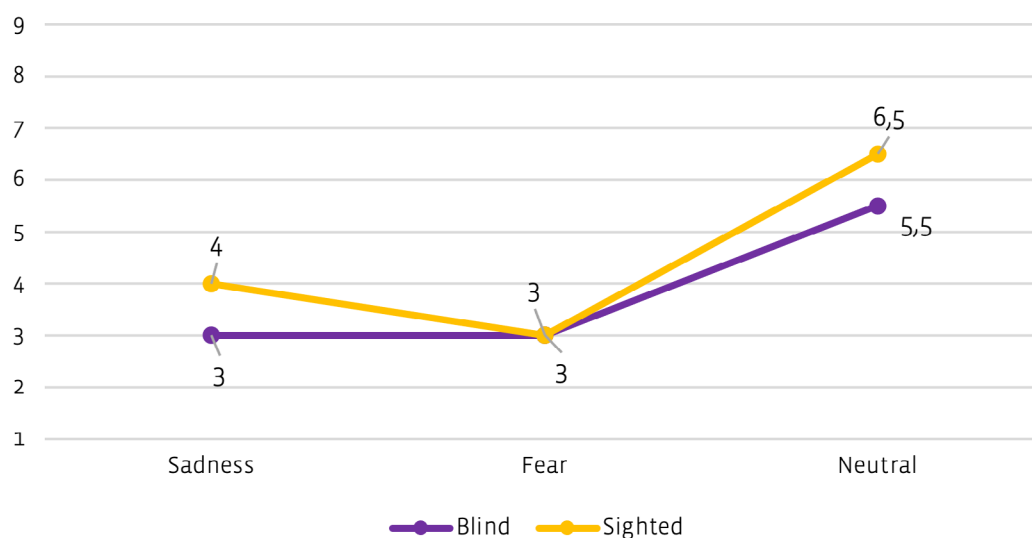
6.1. Self-report measures: T-SAM ratings

In order to compare the ratings of both T-SAM tests, and since the data used are ordinal data, the non-parametric test Mann-Whitney for independent samples was used.

Valence. The analysis yielded significant differences between both groups (Mann-Whitney $U = 1194, p = .004$) (Figure 2). The pairwise comparisons within the Mann-Whitney U test results showed a significant difference between the neutral category values of both groups ($p = .006$).

FIGURE 2

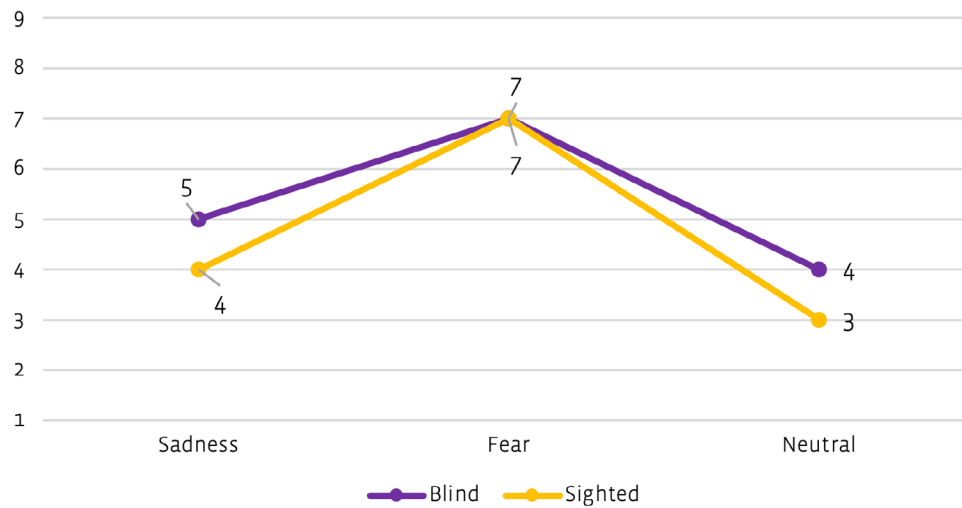
Valence medians for both groups in the three emotions



Arousal. No differences were found comparing the arousal ratings of both groups (Figure 3).

FIGURE 3

Arousal medians for both groups in the three emotions

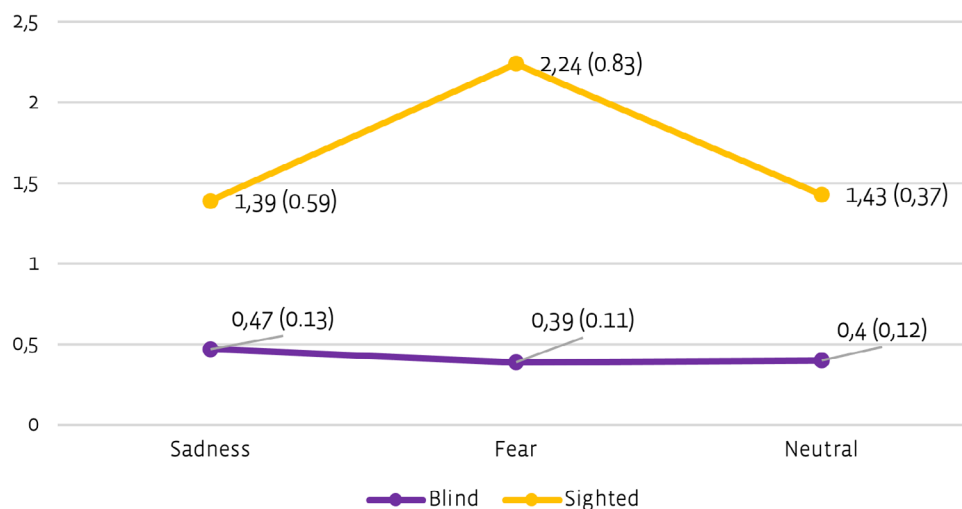


6.2. Psychophysiological measures

EDA. Figure 4 below shows the averages for EDA in both groups across the three emotions. Comparisons between the differences of the average values were performed using a T-test for independent samples. Significant differences were found for neutral ($t_{(67)} = -2.512, p = .014$) and fear ($t_{(67)} = -2,108, p = .039$) categories.

FIGURE 4

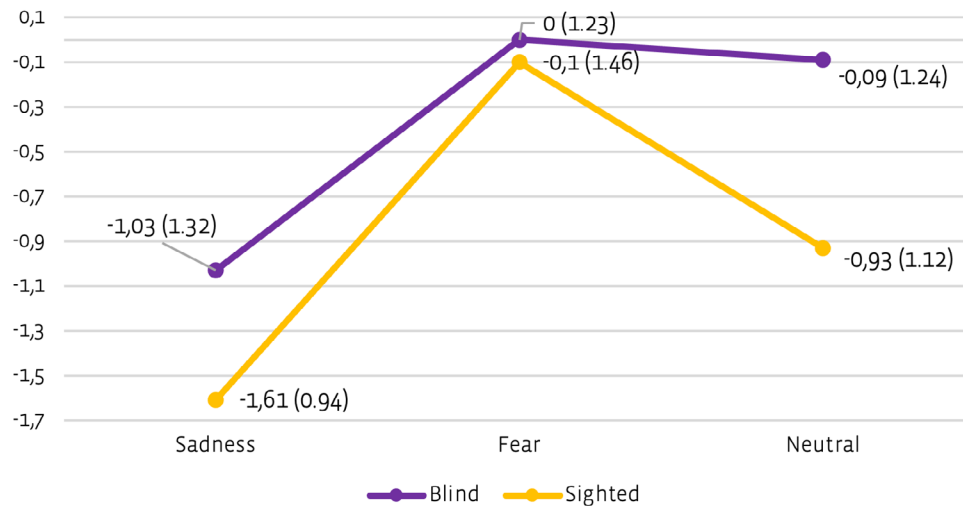
Average values of EDA for blind and sighted participants



HR. Figure 5 below shows averages for the two groups across the two emotions. The T-tests for independent samples performed for the three emotions did not show any significant difference.

FIGURE 5

Average values of HR for blind and sighted participants



6.3. Discussion

In this section, the results of both the subjective measure (SAM) and the psychophysiological measures taken in experiment 1 and 2 are compared. For valence, the statistical analysis yielded some significant differences. The median of the ratings showed that in both groups the structure is maintained: neutral clips were reported to be more positive, followed by sadness and fear, which obtained the lowest valence for both groups. However, sighted participants rated the neutral clip as slightly more positive. For arousal, there were no significant differences between blind and partially sighted and sighted participants. It can be argued that in this case both groups rated the clips following the same pattern and with similar rates.

Regarding EDA results, only significant differences between groups were found for the values obtained during neutral clips, being sighted participants those experimenting higher arousal. On the other hand, HR values did not yield significant differences. As observed in the differences, both groups show the same pattern, in which sadness results in more deceleration, followed by neutral and finally fear, that cause less deceleration.

7. Conclusion

The study of the reception of accessible audiovisual contents is still an under-researched field. However, the massive creation of contents and their growing availability urge not only for the increase

of accessibility services but also for more reception studies. The availability of new technologies also allows for innovative methodologies. Traditional reception studies methods, such as self-report instruments based on scale rating, are combined, as in this case, with physiological measures, deemed more objective. In the field of MA, the application of such methods is still developing and has only been used in a handful of studies (Fryer, 2013; Ramos, 2015). The versatility of this type of measuring tools allows for their application to different fields and with different types of stimuli. In this case, the AST service was compared to the “standard” presentation of written subtitles.

Results have shown that there is similarity in the self-report of emotional arousal of both groups, suggesting that the clips were understood in a similar way in the different treatments of audio-visual translation: audio subtitling and subtitling. In this case, the same can be argued for both dimensions: valence and arousal. According to valence, sadness and fear were rated as more negative than neutral. For arousal, fear was rated higher in comparison to sadness and neutral, reporting less arousal. Thus, both subtitling and audio-subtitling contribute in a similar way to the users’ subjective perception of emotion.

This experiment also points out the effectivity of the questionnaire used, the T-SAM. However, more research is required on the adaptation of questionnaires devoted to the measurement of the emotional state for people who cannot access partly or totally regular instruments used for this purpose (Iturregui-Gallardo & Méndez-Ulrich, 2019).

The results obtained from psychophysiological measures did not provide as clear conclusions as in the self-report. In fact, in studies on media services for the blind and partially sighted (Fryer, 2013; Ramos, 2015), psychophysiological measures did not report solid outcome either. The measurement of physiological markers is far from simple since it is difficult to distinguish which factors had an impact on the recordings. Bodily changes are also dependant on both internal and environmental factors and the oscillations of EDA and HR are directly linked to these. Moreover, while self-report questionnaires elicit a cognitive assessment of the experienced emotion, psychophysiological measures reveal objective changes in body parameters. Participants were not living an emotion (generated by a direct experience) but perceiving and understanding the emotions depicted in the clips. Although the clips are supposed to immerse the spectator in the action, it is possible that the feeling of fear originated by the empathy with the main character in the scene did not activate physiological reactions in the same way a real dangerous situation would do. Therefore, it is worth mentioning that perhaps perceived emotions did not generate bodily reactions strong enough to be recorded.

Further research is needed to assess the psychophysiological effects of emotions depicted in AV contents for both sighted, and blind and partially sighted users. Further exploration of such measuring methods as a tool to evaluate emotional activation is required, and, in any case, it is recommended the combination of psychophysiological measures and self-report instruments in order to provide a better assessment.

8. References

BARTOLL, Eduard, 2006: “Subtitling multilingual films”, paper presented in *MuTra 2006 – Audio-visual Translation Scenarios: Conference Proceedings*, 1-6.

BARTSCH, Anne, Roland MANGOLD, Reinhold VIEHOFF and Peter VORDERER, 2006: “Emotional gratifications during media use - An integrative approach”, *Communications* 31 (3), 261-278.

BENECKE, Bernd, 2012: “Audio description and audio subtitling in a dubbing country: Case studies” in Elisa PEREGO (ed.): *Emerging topics in translation: Audio description*, Trieste: EUT Edizioni Università di Trieste, 99-104.

BRADLEY, Margaret M., Maurizio CODISPOTI, Bruce N. CUTHBERT, and Peter J. LANG, 2001a: “Emotion and Motivation I: Defensive and Appetitive Reactions in Picture Processing”, *Emotion* 1 (3), 276-298.

BRADLEY, Margaret M., Maurizio CODISPOTI, Dean SABATINELLI, and Peter J. LANG, 2001b: “Emotion and Motivation II: Sex Differences in Picture Processing”, *Emotion* 1 (3), 300-319.

BRADLEY, Margaret M., and Peter J. LANG, 1994: “Measuring Emotion: The Self-Assessment Semantic Differential Manikin and the Semantic Differential”, *Journal of Behavior Therapy and Experimental Psychiatry* 25 (1), 49-59.

BRAUN, Sabine, and Pilar ORERO, 2010: “Audio description with audio subtitling – an emergent modality of audiovisual localisation”, *Perspectives: Studies in Translatology* 18 (3), 173-188.

CHAUME, Frederic, 2003: *Doblatge i subtitulació per la TV*, Vic: Eumo.

CHAUME, Frederic, 2010: “Traducción audiovisual” in Luis PEGENAUTE and FRANCISCO LAFARGA (eds.): *Diccionario Histórico de la Traducción en España*, Madrid: Gredos.

CSIKSZENTMIHALYI, Mihaly, 1997: *Finding flow: The psychology of engagement with everyday life*, New York: Basic Books.

DÍAZ CINTAS, Jorge, 1999: “Dubbing or subtitling: The eternal dilemma”, *Perspectives: Studies in Translatology* 7 (1), 31-40.

DÍAZ CINTAS, Jorge, 2001: *La traducción audiovisual. El subtitulado*, Salamanca: Almar.

DÍAZ CINTAS, Jorge, and Aline REMAEL, 2007: *Audiovisual Translation: Subtitling*, Manchester: St. Jerome.

DILLON, Cath, 2006: *Emotional responses to immersive media*. PhD Thesis, University of London.

DILLON, Cath, Edmund KEOGH, Jonathan FREEMAN, and Jules DAVIDOFF, 2001: “Presence: is your heart in it”, paper presented in the 4th Annual International Workshop on Presence.

DRACHEN, Anders, Lennart E. NACKE, Georgios YANNAKAKIS, and Anja Lee PEDERSEN, 2010: “Correlation between heart rate, electrodermal activity and player experience in first-person shooter games”, paper presented in the 5th ACM SIGGRAPH Symposium on Video Games - Sandbox '10.

EARDLEY-WEAVER, Sarah, 2014: *Lifting the Curtain on Opera Translation and Accessibility: Translating Opera for Audiences with Varying Sensory Ability*. PhD Thesis, Durham University.

EGAN, Darragh, Sean BRENNAN, John BARRETT, Yuansong QIAO, Christian TIMMERER, and Murray NIAIL, 2016: “An evaluation of Heart Rate and ElectroDermal Activity as an objective QoE evaluation method for immersive virtual reality environments”, paper presented in the 2016 Eighth International Conference on Quality of Multimedia Experience (QoMEX), 1-6.

FRYER, Louise, 2013: *Putting It Into Words: The Impact of Visual Impairment on Perception, Experience and Presence*. PhD Thesis, University of London.

GAMBIER, Yves, 2006: “Multimodality and audiovisual translation”, paper presented in MuTra2006 – Audio visual Translation scenarios.

GRAHAM, Frances K., 1992: “Attention: The Heartbeat, the Blink, and the Brain” in Byron A. CAMPBELL, Harlene HAYNE, and Rick RICHARDSON (eds.): *Attention and Information Processing in Infants and Adults*, New York: Psychology Press.

GROSS, James J., and LEVENSON, Robert W., 1995: “Emotion elicitation using films”, *Cognition & Emotion* 9 (1), 87-108.

ITURREGUI-GALLARDO, Gonzalo, 2018: “Rendering multilingualism through audio subtitles: shaping a categorisation for aural strategies”, *International Journal of Multilingualism*, 1-14.

ITURREGUI-GALLARDO, Gonzalo, Anna MATAMALA, and Olga SOLER-VILAGELIU, 2018: “Audio Subtitling : Measuring Emotional Arousal in Users Through Psychophysiological Measures”, paper presented in the Languages & the Media Conference.

ITURREGUI-GALLARDO, Gonzalo, and Jorge-Luís MÉNDEZ-ULRICH, 2019: “Towards the Creation of a Tactile Version of the Self-Assessment Manikin (T-SAM) for the Emotional Assessment of Visually Impaired People”, *International Journal of Disability, Development and Education* 67 (6), 657-674.

KETTUNEN, Joni, Niklas RAVAJA, Petri NÄÄTÄNEN, Pertti KESKIVAARA and Liisa KELTIKANGAS-JÄRVINEN, 1998: “The synchronization of electrodermal activity and heart rate and its relationship to energetic arousal: a time series approach”, *Biological Psychology* 48 (3), 209-225.

KLEINGINNA, Paul R., and Anne M. KLEINGINNA, 1981: “A categorized list of emotion definitions, with suggestions for a consensual definition”, *Motivation and Emotion* 5 (4), 345-379.

KLIMMT, Christoph, and Peter VORDERER, 2003: “Media Psychology ‘is not yet there’: Introducing Theories on Media Entertainment to the Presence Debate”, *Presence: Teleoperators and Virtual Environments* 12 (4), 346-359.

KREIBIG, Sylvia D., 2010: “Autonomic nervous system activity in emotion: A review”, *Biological Psychology* 84 (3), 394-421.

KREIBIG, Sylvia D., Frank H. WILHELM, Walton T. ROTH, and James J. GROSS, 2007: “Cardiovascular, electrodermal, and respiratory response patterns to fear- and sadness-inducing films”, *Psychophysiology* 44 (5), 787.

KRUGER, Jan Louis, Maria Teresa SOTO-SANFIEL, Stephen DOHERTY, and Ronny IBRAHIM, 2016: “Towards a cognitive audiovisual translatology” in Ricardo MUÑOZ MARTÍN (ed.): *Reembedding Translation Process Research*, Amsterdam: John Benjamins, 171-194.

LANG, Peter J., Margaret M. BRADLEY, and Bruce N. CUTHBERT, 2008: *International Affective Picture System (IAPs): Affective ratings of picture and instruction manual (Technical Report A-8)*, Gainesville: University of Florida.

LJUNGLÖF, Peter, Sandra DERBRING, and Maria OLSSON, 2012: “A Free and Open-Source Tool That Reads Movie Subtitles Aloud”, paper presented in the *Workshop on Speech and Language Processing for Assistive Technologies*, 1-4.

MANDRYK, Regan L., Kori M. INKPEN, and Thomas W. CALVERT, 2006: “Using psychophysiological techniques to measure user experience with entertainment technologies”, *Behaviour & Information Technology* 25 (2), 141-158.

MATAMALA, Anna, Elisa PEREGO, and Sara BOTTIROLI, 2017: “Dubbing versus subtitling yet again? An empirical study on user comprehension”, *Babel* 63 (3).

MAĆCZYŃSKA, Magdalena (2011). *Text-to-speech audio description with audio subtitling to a non-fiction film. A case study based on La Soufriere by Werner Herzog*. MA Dissertation, University of Warsaw.

McMANIS, Mark H., Margaret M. BRADLEY, Keith W. BERG, Bruce N. CUTHBERT, and Peter J. LANG, 2001: “Emotional reactions in children: Verbal, physiological, and behavioral responses to affective pictures”, *Psychophysiology* 38, 222-231.

MIESENBERGER, Klaus, Joachim KLAUS, and Wolfgang ZAGLER, 2002: “Computers helping people with special needs”, paper presented in the *8th International Conference, ICCHP 2002*, Linz: Springer.

MIHKLA, Meelis, Indrek HEIN, Indrek KIISSEL, Artur RÄPP, Risto SIRTS, and Tanel VALDNA, 2014: “A System of Spoken Subtitles for Estonian Television”, *Human Language Technologies - the Baltic Perspective* 268, 19-26.

MIKROPOULOS, Tassos A., Evangelos TZIMAS, and Georgios EL. DIMOU, 2004: “Objective Presence Measures through Electric Brain Activity”, paper presented in the *Proceedings of the 7th Annual International Workshop on Presence: Presence 2004* (February), 259-265.

MONTEFINESE, Maria, Ettore AMBROSINI, Beth FAIRFIELD, and Nicola MAMMARELLA, 2014: “The adaptation of the Affective Norms for English Words (ANEW) for Italian”, *Behavior Research Methods* 46 (3), 887-903.

MOVIETALK, 2014: *Spoken Subtitles and Audio Description in Cinema. Zero Project.*

NIELSEN, Simon, and Hans-Heinrich BOTHE, 2008: “Subpal: A device for reading aloud subtitles from television and cinema”, paper presented in the *CEUR Workshop Proceedings* (vol. 415).

NOWLIS, Vincent, 1965: “Research with the Mood Adjective Check List” in Silvan S. TOMPKINS and Carroll E. IZARD (eds.): *Affect, cognition, and personality: Empirical studies*, Oxford: Springer, 464.

O’HAGAN, Minako, 2010: “Engineering Emotion? Translators’ New Role in Serving the Digital Entertainment Industry” (Keynote lecture), *International Conference on Translation and Accessibility in Video Games and Virtual Worlds*, Barcelona: Autonomous University of Barcelona.

O’HAGAN, Minako, 2016: “Game Localisation as Emotion Engineering: Methodological Exploration” in Qi ZHANG and Minako O’HAGAN (eds.): *Conflict and Communication: A Changing Asia in a Globalising World*, New York: Nova, 123-144.

ORERO, Pilar, 2007: “Audiosubtitling: a possible solution for opera accessibility in Catalonia”, *Tradterm* 13, 135.

ORERO, Pilar, 2011: “The audiodescription of spoken, tactile and written language in Be with me” in Anna MATAMALA, Jean-Marc LAVAU, and Adriana SERBAN (eds.): *Audiovisual Translation in Close-up*, Bern: Peter Lang, 239-255.

PEREGO, Elisa, Fabio DEL MISSIER, and Sara BOTTIROL, 2014: “Dubbing versus subtitling in young and older adults: cognitive and evaluative aspects”, *Perspectives* 6623 (July), 1-21.

PEREGO, Elisa, Fabio DEL MISSIER, Marco PORTA, and Mauro MOSCONI, 2010: “The Cognitive Effectiveness of Subtitle Processing”, *Media Psychology* 13 (3), 243-272.

PETRIE THOMAS, Julianne H., Michael F. WHITFIELD, Tim F. Oberlander, Anne R. SYNNEs, and Ruth E. GRUNAU, 2012: “Focused attention, heart rate deceleration, and cognitive development in preterm and full-term infants”, *Developmental Psychobiology* 54 (4), 383-400.

RAMOS CARO, Marina, 2016: *La traducción de los sentidos*, Munich: Lincom GmbH.

RAMOS CARO, Marina, 2015: “The emotional experience of films: Does audio description make a difference?” *Translator* 21 (1), 68-94.

REEVES, Byron, Annie LANG, Eun Young KIM, and Deborah TATAR, 2009: “The Effects of Screen Size and Message Content on Attention and Arousal”, *Media Psychology* 1 (1), 49-67.

REMAEL, Aline, 2012: “Audio Description with Audio Subtitling for Dutch Multilingual Films: Manipulating Textual Cohesion on Different Levels”, *Meta: Journal Des Traducteurs* 57 (2), 385-407.

REVIERS, Nina, and Aline REMAEL, 2015: “Recreating multimodal cohesion in audio description: A case study of audio subtitling in Dutch multilingual films”, *New Voices in Translation Studies* 13 (1), 50-78.

ROONEY, Brendan, Claran BENSON, and Ellis HENNESSY, 2012: “The apparent reality of movies and emotional arousal: A study using physiological and self-report measures”, *Poetics* 40 (5), 405-422.

SEPIELAK, Katarzina, 2016: “The Effect of Subtitling and Voice-Over on Content Comprehension and Language Identification in Multilingual Movies”, *International Journal of Sciences: Basic and Applied Research* 25 (1), 166-182.

SHERRY, John L., 2004: “Flow and Media Enjoyment”, *Communication Theory* 14 (4), 328-347.

SIMONS, Robert F., Benjamin H. DETENBER, and Thomas M. ROEDEMA, 1999: “Emotion processing in three systems: The medium and the message”, *Psychophysiology* 36, 619-627.

TAYLOR, Shelley E., 1991: “Asymmetrical effects of positive and negative events: The mobilization-minimization hypothesis”, *Psychological Bulletin* 110 (1), 67-85.

THRANE, Lisbeth Kvistholm, 2013: *Text-to-Speech on Digital TV: An Explanatory Study of Spoken Subtitles on DR1Syn*. MA dissertation, IT University of Copenhagen.

TREMAYNE, Patsy, and Robert J. BARRY, 2001: “Elite pistol shooters: physiological patterning of best vs. worst shots”, *International Journal of Psychophysiology* 41 (1), 19-29.

UDO, John Patrick, Bertha ACEVEDO, and Deborah I. FELS, 2010: “Horatio audio-describes Shakespeare’s Hamlet: Blind and low-vision theatre-goers evaluate an unconventional audio description strategy”, *British Journal of Visual Impairment* 28 (2), 139-156.

UHRIG, Meike K., Nadine TRAUTMANN, Ulf BAUMGÄRTNER, Rolf-Detlef TREEDE, Florian HENRICH, Wolfgang HILLER, and Susanne MARSCHALL, 2016: “Emotion Elicitation: A Comparison of Pictures and Films”, *Frontiers in Psychology* 7 (February).

WALCZAK, Agnieszka, 2017: *Immersion in audio description*. Unpublished PhD Thesis, Autonomous University of Barcelona.

WALCZAK, Agnieszka, and Louise FRYER, 2018: “Vocal delivery of audio description by genre: measuring users’ presence”, *Perspectives* 26 (1), 69-83.

WATSON, David, and Lee A. CLARK, 1999: *The PANAS-X: Manual for the Positive and Negative Affect Schedule - Expanded Form*, Iowa City: Department of Psychological and Brain Sciences Publications.

WESTERMANN, Rainer, Kordelia SPIES, Günther STAHL, and Friedrich W. HESSE, 1996: “Relative effectiveness and validity of mood induction procedures: A meta-analysis”, *European Journal of Social Psychology* 26 (4), 557-580.

WILKEN, Nicole, and Jan-Louis KRUGER, 2016: “Putting the audience in the picture: Mise-en-shot and psychological immersion in audio described film”, *Across Languages and Cultures* 17 (2), 251-270.

WILSON, Gillian M., and Angela M. SASSE, 2006: “Do Users Always Know What’s Good For Them? Utilising Physiological Responses to Assess Media Quality” in Sharon McDONALD, Yvonne WAERN, and Gilbert COCKTON (eds.): *People and Computers XIV — Usability or Else! Proceedings of HCI 2000*, London: Springer London, 327-339.