Audio subtitling:

Dubbing and voice-over effects and their impact on user experience

Iturregui-Gallardo, Gonzalo

gonzalo.iturregui@uab.cat

Matamala, Anna

anna.matamala@uab.cat

Department of Translation and Interpreting & East Asian Studies, Universitat Autònoma de

Barcelona, Spain

Abstract

Although there has been recent research on other media accessibility services such as audio description, there has been little focus on audio subtitling and the way subtitles are delivered orally. This article reports the outcome of an experiment in which 42 Spanish blind and partially sighted participants were exposed to two diverging audio subtitling strategies: audio subtitles with a voice-over effect and audio subtitles with a dubbing effect. Data on the users' emotional responses were collected through a tactile and simplified version of the SAM questionnaire and psychophysiological recordings of electrodermal activity and heart rate. The results obtained from both methods do not show statistically significant differences between the two effects. However, results from the questionnaire proved that emotions were induced in the participants calling for more research on the topic and with the application of such methods.

1. Introduction

This article is framed within a tradition of studies on audiovisual translation and media accessibility that investigate user experience. Extensive research on the reception of the main transfer modes in audiovisual translation, such as dubbing, subtitling, and voice-over, has been carried out in recent decades. To name just a few, they include Matamala et al. (2017) and Perego et al. (2014) for dubbing and subtitling; Perego et al. (2010), Kruger et al. (2016) for subtitling, and Sepielak (2016) for subtitling and voice-over.

In the field of media accessibility, subtitling for the Deaf and Hard-of-Hearing has attracted considerable attention (Romero-Fresco, 2016) and there is also extensive research on audio description (AD). The reception of AD has been studied when applied to the theatre (Udo et al., 2010), to fashion shows (Udo & Fels, 2009), opera (Cabeza-Cáceres, 2010; Cabeza-Cáceres & Matamala, 2008; Eardley-Weaver, 2014; Matamala & Orero, 2007; Orero, 2007; Orero & Matamala, 2007) and cinema (Fresno Cañada, 2014; Cabeza-Cáceres, 2013; Fernández-Torné & Matamala, 2016; Szarkowska, 2011). Studies in this field have focused on many aspects, such as information recall (Fresno Cañada, 2014) or user comprehension and enjoyment (Cabeza-Cáceres, 2013) when faced with diverging AD strategies. The reception of synthetic as opposed to human voices has also been the subject of research (Fernández-Torné & Matamala, 2016; Szarkowska, 2011). The methods used in these investigations have generally been based on self-report questionnaires (Fresno Cañada, 2014), tools such as eye-tracking (Di Giovanni, 2014) as well as some standardised tests (Fryer & Freeman, 2014).

The present article focuses on the study of audio subtitling, which allows to access written subtitles in their aural form. More specifically, it aims to take a step forward by researching two strategies for the delivery of audio subtitles and their impact on the emotional reaction of persons with sight loss. The two strategies under analysis are the so-called "dubbing effect" and "voice-over effect" (see section 3). In line with recent studies (O'Hagan,

2016), the experiment uses innovative methods in the field of media accessibility based on physiological reactions (Fryer, 2013; Ramos Caro, 2016; Ramos, 2015): electrodermal activity (EDA) and heart rate (HR).

Overall, the main objective of the experiment presented in this article is to compare two audio subtitling delivery effects (dubbing effect/voice-over effect) in terms of the emotional arousal in persons with sight loss in Spain. The main hypothesis is that a dubbing effect will result in higher reported emotional arousal and also in higher values in the psychophysiological measures used (EDA and HR). This hypothesis is based on the fact that the targeted participants should live in an audiovisual environment with a strong dubbing tradition.

The article is structured as follows: Section 2 discusses the concept of emotion and establishes the relationship between the study of emotional arousal and psychophysiology. Section 3 defines audio subtitling and the two strategies under analysis. Section 4 describes the methodological framework, and section 5 reports on the results and discusses them. In section 6, conclusions and further research avenues complete the article.

2. Emotional arousal and psychophysiological measures

Research on emotion has increased in recent decades (Rottenberg et al., 2007). Different terms have been used to refer to the study of emotions, which reflect diverging approaches: emotionology (Stearns & Stearns, 1985), emotioncy (Pishghadam, Firooziyan, & Esfahani, 2017), affective neuroscience (Lindquist, Kober, & Barrett, 2015) or affective science (Rottenberg, 2003) are examples of such terminological variation. The definition of emotion itself has also been proposed in different manners in the field (Izard, 2010; Kreibig, 2010; Plutchik, 2009; Rottenberg et al., 2007; Scherer, 2005), yet such definitions tend to share similar ideas. Based on the previous authors, an emotion is understood in this paper as the response of the organism, at the behavioural, social and neurophysiological levels,

provoked by a single stimulus or a series of stimuli that can be both internal or external. The stimuli are processed by our cognitive systems, which are responsible for distributing the information that generates emotional arousal.

Different methods are used for the assessment of emotional arousal, which can be defined as the activation of the aforementioned responses in the organism. Mauss & Robinson (2009, p. 14) point out that there is no "gold standard" for their measurement but the more measures that are applied the more that can be learnt about emotional arousal caused by certain stimuli (Larsen & Prizmic-Larsen, 2006). However, there are two main types of methods (Lang 1969): affective reports and physiological changes. A high number of questionnaires have been designed to assess the emotional state such as 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). Although useful, these questionnaires are regarded as subjective in nature and are generally used in combination with physiological data. Changes in the physiology of the organism have proved to correlate with what we feel and experience (Kreibig, 2010) and they are captured through biofeedback technology, by recording indicators of emotional arousal such as brainwaves, muscle contraction, electrodermal activity (EDA) or heart rate (HR).

The measurement of emotional reaction when users watch audiovisual content has attracted the interest of many researchers (Uhrig et al., 2016). To name just a few, O'Hagan (2016) analysed physiological data from HR and EDA to measure the gamer emotions derived from playing a videogame. Wilson & Sasse (2006) used HR, EDA and blood volume pulse when testing the quality of the image based on frame segmentation in video stimuli. Physiological responses were used to test interactive technologies in Mandryk, Inkpen, & Calvert (2006). Moving on to the measurement of the response generated by cinema, Dillon (2006) collected data from EDA and HR when exposing participants to emotional and non-emotional films. More recently, electroencephalography (EEG) was used in a study by Kruger

et al. (2016) on immersion in subtitled audiovisual material. In terms of media accessibility, to the best of our knowledge, the impact of AD on blind and partially-sighted audiences has only been studied through the use of physiological markers in two studies. Heart rate was used by Ramos (2015) in combination with the PANAS-X questionnaire (Watson & Clark, 1999) to test the emotional arousal of participants when exposed to audio described emotional clips; and HR and EDA was used by Fryer (2013) in combination with questionnaires on presence (the ITC-SOPI questionnaire) to test the immersion of different types of audio descriptions in clips eliciting three targeted emotions: amusement, fear, and sadness.

3. Audio subtitling: dubbing and voice-over effect

Audio subtitling could be summed up as a spoken or aural rendering of subtitles. This is the approach taken by authors such as Nielsen & Bothe (2008), Miesenberger, Klaus, & Zagler (2002, p. 296), Fryer (2013: 194), Szarkowska & Mączyńska (2011, p. 24), Orero (2011, p. 237), the International Organization for Standardization & International Electrotechnical Commission (2015) or the International Telecommunication Union (2013, p. 3). However, a definition that most clearly illustrates the heterogeneity of this service is provided by Reviers & Remael (2015, p. 52):

> AST can therefore be defined as the aurally rendered and recorded version of the subtitles with a film. This spoken version of the subtitles is mixed with the original soundtrack. AST is usually read, sometimes acted out, by one or more voice actors. Sometimes it is produced by text-to-speech software. The subtitle text is often delivered almost literally but it can be rewritten to varying degrees and, in addition, the recording method also varies. Usually, AST is recorded as a form of voice-over, which means that the original dialogues can be heard briefly before the translation starts. Sometimes it is recorded in a semi-dubbed form, which means that the original dialogues are substituted by a form of dubbing that

is not necessarily entirely lip-sync, that is, synchronous with the lip movement of the speaker.

This access service is placed somewhere between the adaptation of audiovisual contents for people with specific needs and the more canonical concept of translation, which is based on the translation proper (interlingual translation), as expressed by Jakobson (2000/1959), where a message in a source language (SL) is transferred into a target language (TL). The SL is translated into a TL in the written subtitles and then delivered orally in the same TL.

Audio subtitles can be found as an independent access service or in combination with audio description. As an independent service, some countries have developed systems that allow for written subtitles to be read aloud with text to speech technologies (Mihkla et al., 2014; Thrane, 2013). However, with the expansion of digital television and on-demand services, devices such as smartphones allow for the reading-aloud of subtitles without the need for specific software (Royal National Institute of Blind People, 2017). On the other hand, as part of the AD track, audio subtitles are voiced by voice talents or describers themselves, creating a single unit in which the combination of the two services exists from the beginning of the audio description process. In this case, the AD and AST parts are differentiated by using different voices or by using a single voice with different intonations.

The way in which audio subtitles are delivered provides different information about the message or the multilingual reality of the audiovisual content in which they are used. As Iturregui-Gallardo (2018) suggests, by applying variations in aspects such as the volume, the acting or the information provided by the audio describer, audio subtitles can take on different forms. Two of the main strategies in terms of voicing, which have already been discussed in previous works (Braun & Orero, 2010; ISO/IEC 20071-25, 2017; Remael, Reviers, & Vercauteren, 2014), are dubbing and voice-over. These strategies are also known under the name of "effects", since they provide an effect similar to that of these two transfer modes.

According to the aforementioned works, these two AST effects are mainly differentiated by three features: (1) volume of the original and the audio subtitle tracks, (2) isochrony between the original utterance and its corresponding audio subtitle, and (3) prosodic features in the delivery of audio subtitles. Based on the existing literature on the topic and the definition of the two transfer modes, the AST effects can be described as follows:

(1) Voice-over effect: in which the AST is displayed a short time after the original, which can be heard underneath. The voice used can be acted to some extent but its prosodic features are related to reading-aloud, which is closer to the written code.

(2) Dubbing effect: in which the AST is displayed in synchrony with the original line, which cannot be heard. The voice used is acted, in the sense that it replicates the emotive tone in the dialogue, similar to the standardised prefabricated orality attributed to dubbing.

4. Method

The study combined self-reporting and psychophysiological measures to test the emotional arousal experienced by Spanish blind and partially-sighted participants when exposed to Polish clips with audio subtitles with a dubbing and a voice-over effect. This section describes the participants, instruments, stimuli used and experimental procedure.

4.1. Participants

The main selection criterion was that participants should not have access to the written subtitles and that they should be Spanish native speakers from Spain, a traditionally dubbing country. No further selection criteria were set, as the aim was to gather a diverse sample of users. To recruit participants, a Barcelona-based association of persons who are blind or

partially sighted (Associació Discapacitat Visual Catalunya - B1B2B3) was contacted. 42 participants took part in the experiment, a higher number than normal in media accessibility research (Orero et al., 2018). The experiment was carried out with 13 blind and 29 partially sighted participants (N = 42). There were 17 women and 25 men of whom 19 had finished university studies, 14 had finished professional training degree, 5 had finished secondary school, 2 had finished primary school and 1 did not attend school. They were aged between 19 and 73 years old (median = 38). No further demographic information was gathered. Each participant was asked to complete a consent form, that was read out loud to them. The experiment and the forms to be used were approved by the ethics committee of the Universitat Autònoma de Barcelona.

4.2. Instruments and measures

The experiment combined a self-report questionnaire with psychophysiological data. The Tactile Self-Assessment Manikin (T-SAM) was used together with EDA and HR measures, as it was considered that a combination of objective and subjective measures would contribute to a better understanding of emotional activation.

4.2.1. Self-report questionnaire

The T-SAM questionnaire consists of a tactile, simplified and augmented version of the SAM questionnaire. The original questionnaire designed in the 1990s by Bradley & Lang (1994) is based on the emotional model postulated by Lang (1988) and argues that the emotional experience is composed of three fundamental dimensions: the emotional valence (the emotional value, more or less positive or negative, attributed to the situation or stimulus that triggers the emotion), the arousal or perception of physiological activation induced by the emotion (from not at all activating to extremely activating), and the dimension of dominance or feeling of control experienced over emotion (from no control to complete control).

The questionnaire is made up of three sub-scales that evaluate each of the components of the emotions according to Lang's model (1988). In these scales, an illustration represents different levels (from minimum to maximum) of emotional valence, arousal, and dominance. It is accompanied by a numerical 9-point scale that participants have to complete by indicating their rating on each one of these scales at the precise moment when they take the test. Due to its "brevity, simplicity and transcultural character (its dimensions are universal and not susceptible to contamination from cultural values and patterns)" (Iturregui-Gallardo & Méndez-Ulrich, 2019, p. 3), the SAM represents one of the gold standards in emotional evaluation, especially in experimental situations of emotional induction, and has been used widely in many studies (Bradley, Codispot, Chuhbert, & Lang, 2001; Backs, da Silva, & Han, 2005; Betella & Verschure, 2016; Geethanjali, Adalarasu, Hemapraba, Pravin Kumar, & Rajasekeran, 2017).

For the adapted version of the SAM, the third dimension of dominance has been removed, as recommended in more recent studies (Lang, Bradley, & Cuthbert, 2008), as it is not clear if this dimension differs from that of valence (Montefinese, Ambrosini, Fairfield, & Mammarella, 2014) and is less related to physiological arousal. The experiments combined the questionnaire with psychophysiological measures that are linked to its two first scales as they aim at the measurement of emotional arousal. It was then modified to a simplified, augmented and tactile version that was designed to suit participants with different kinds of requirements. Two proposals with different design options were vectorized and then printed with UV gloss on a non-absorbent surface (PVC). The final version (Figure 1) was created based on the input received from blind and partially sighted persons in a focus group (Iturregui-Gallardo & Méndez-Ulrich, in press).

Figure 1. Image of the Tactile-Self Assessment Manikin questionnaire

4.2.2. Psychophysiology

Two different psychophysiological markers related to emotional reactions were recorded in this experiment: EDA and HR. Electrodermal activity measures the electrical conductance of the skin in microsiemens whilst heart rate measures beats per second. The combination of EDA and HR is commonly used, as the tools are relatively cheap and available and their application is not invasive for the participant (Matamala et al., in press). Furthermore, Kettunen et al. (1998) stated that there is synchronisation between EDA and HR and, in their study, both measures correlated with self-report measures. The CAPTIV L7000 device and software were used. This is consists of a central device, which is connected to a computer and receives wireless data from the sensors. Two sensors were used in this case: the first one, for EDA recording, is divided into two sensors that are attached to the skin under the second phalange of the index and middle fingers; the second one, for HR recording, is shaped like a belt and attached to the skin at the end of the sternum. Data were recorded at a frequency of 32 Hz. The CAPTIV software was linked to Tobii Studio software, which provided a simple tool for procedure set-up, worked as an experimental controller and automated the presentation of stimuli and instructions to the participants.

4.3. Stimuli

The criteria for the selection of the video excerpts were as follows: the excerpts should be in a language which the participants could not understand so that they would need to rely on the audio subtitles in Spanish; they should be self-contained and have a length of about 3 minutes. The duration of the stimuli in this type of research varies within the scientific community. Using clips that last from some seconds to more than 10 minutes has proven to cause physiological activation (Kreibig et al., 2007; Rottenberg et al., 2007; Rooney et al., 2012). It was also decided that they should display negative emotions such as anger, fear or sadness and that a neutral clip with no emotions should also be included. Negative emotions were selected because they are easier to induce, create stronger reactions (Uhrig et al., 2016;

Westermann, Spies, Stahl, & Hesse, 1996) and are the ones most frequently used in emotion research (Kreibig, 2010). The inclusion of a neutral clip had been suggested in previous studies (Kreibig, Wilhelm, Roth, & Gross, 2007). It was decided that the selection of the audiovisual stimuli would be carried out by means of an online validation to avoid subjective bias on the part of the researcher.

Eight scenes from the Polish TV series *Wojenne Dziewczyny* [War Girls][1] (2017), broadcast by the Polish national channel *Telewizja Polska 1* (TVP1) were selected by the researcher and permission to use the clips was obtained. The use of a single source guaranteed cohesion in terms of acting, character voices, plot, and musical and ambience effects. These 8 scenes, which were then subtitled into Spanish, were selected to correspond to the emotions of anger (2), fear (2), sadness (2) and neutral emotion (2).

In order to confirm that these scenes actually transmit the emotions identified by the researchers, an online validation by means of a survey was carried out with 120 participants. They were requested to watch the eight selected scenes and reply to (a) a question about the emotion they experienced during the viewing, and (b) the first two parts of the previously mentioned SAM questionnaire. From the eight clips, one was chosen for each emotion (sadness, fear, anger, and neutral), i.e. the one with the highest percentage of respondents identifying the emotion and with more accurate valence and arousal ratings according to the literature. At this stage, there were 4 clips. After taking into account that anger and fear are emotions that can easily be confused when analysing their psychophysiological effects, the one with higher scores was chosen. At the end of this validation process there were three clips transmitting fear, sadness, and a neutral emotion on the online validation.

These three clips were audio described by a professional describer in Spanish and, for each of the clips, two versions with audio subtitles were recorded by professional voice talents in a dubbing studio: one with audio subtitles with a dubbing effect and one with audio

---

[1] https://vod.tvp.pl/website/wojenne-dziewczyny,28767487

subtitles with a voice-over effect, as these are the effects that were to be compared in the experiment. The final result was six clips, as shown in Table 1: sadness with voice-over AST, sadness with dubbing AST, fear with voice-over AST, fear with dubbing AST, neutral with voice-over AST, and neutral with dubbing AST.

Table 1. Stimuli of the experiments

4.4. Procedure

Participants were individually welcomed in a controlled environment and were informed about the experiment. They were required to give their informed consent. Then they were asked some preliminary demographic questions. The sensors were attached, the set tested and then instructions were given verbally to participants, including an introduction to the TV series and the plot.

Each participant watched three clips. The presentation of stimuli was randomised across the participants following a between-subject design. All participants watched both AST effects in different clips. This process was achieved by creating 6 different protocols resulting from the combination of the three clips and the two AST conditions, and automatically randomising this across the 42 participants, so that all the combinations were repeated 7 times. This meant that all the stimuli were watched by the same number of participants.

Before the presentation of the first clip the participant was induced into an 8-minute relaxation period with relaxing music. After the sound of a bell, which warned participants that the relaxation period was over, a short description of the scene was provided (see Annex). After each of the clips, participants were asked to respond to the T-SAM questionnaire and a relaxation period of 3 minutes without music followed. At the end a series of preference questions were asked. They are not reported in this article, as the focus is on the data obtained through the measurement of emotional arousal.

5. Discussion of results

The discussion of results is presented in two separate sections. The first presents the outcome and following discussion for the T-SAM questionnaire, more specifically on valence and arousal (see 4.2.1.), which as explained above are the two first dimensions in the SAM questionnaire. Valence rates the emotion in terms of being positive or negative, and arousal rates the participant's level of activation by the emotion. The second section focuses on the values obtained from the psychophysiological measurements, specifically EDA and HR (see 4.2.2.). Statistical analyses were carried out on both the rates obtained in the self-report questionnaire and the values obtained through EDA and HR. More specifically, ANOVA was considered to be suitable to explore the differences in the averages for three different variables. Spearman Rank Correlations were run to assess the correlations: this test allowed for the analysis of the relationship between the psychophysiological results and the dimension of arousal in the self-report questionnaire, which focused on the participant being excited or calmed, and also between the two psychophysiological measures.

5.1. T-SAM questionnaire

Participants gave a rating for each of the two dimensions in the questionnaire: valence and arousal. The results for the valence and arousal ratings for the AST effect were obtained by calculating the average of the ratings of the 42 participants in the experiment. The presentation of the results follows the same structure.

Valence. Figure 2 shows the average results for the valence dimension for the two AST effects (dubbing and voice-over, see section 3). The ANOVA shows no statistically significant differences for the participants' valence ratings on the way the audio subtitles were presented ($F_{(1, 120)} = 0.918$; $p = .34$).

Figure 2. Valence average ratings for AST effect

Figure 3 below shows the average results for the valence dimension for both conditions.

Figure 3. Valence average ratings for AST effect and emotion

It could be argued that the intended emotions were induced during the procedure by the stimuli used and the expected valence ratings were obtained. The difference between the ratings for the emotions (fear, sadness, neutral) of the clips is significant ($p < .001$). However, the audio effect (dubbing and voice-over) proves to be non-significant ($p = .34$) when observed overall for the three emotions.

The only significant difference observed is between emotions, irrespective of the AST effect ($p < .001$), which suggests that they were successfully induced in blind and partially sighted participants according to the T-SAM self-report instrument.

Arousal. Figure 4 shows the averages for the arousal dimension, taking the two AST effects into account. The ANOVA shows that there are no significant differences between the two AST effects ($F (1, 120) = 0.183$; $p = .67$).

Figure 4. Arousal average ratings for AST effect

Figure 5 shows the interaction between the average rates for the conditions of emotion and AST effect. This interaction does not show statistically significant differences ($F (2, 120) = 3.479$; $p = .39$) when applying ANOVA. The relationship between the intended and the rated emotions was very coherent and statistically significant ($p < .001$). The fear-inducing scene is the one that has the highest reported arousal, followed by the sadness-inducing clip

and, finally, the neutral clip. This shows that the emotions were induced. However, the AST effect did not show any statistic difference in arousal ($p = .67$).

Figure 5. Arousal average ratings for emotion and effect

The averages of the ratings for arousal taking into consideration the AST effect are not statistically significant ($p = .67$). However, the results seem similar to the valence, in that clips with a dubbing effect obtain a higher rating (5.46) compared with the voice-over effect (5.27).

The interaction between the conditions of emotion and AST effect for the dimension of arousal did not show any significant difference (F (2, 120) = 3.479; $p = .39$). In the overall pattern, the most noteworthy difference observed between AST effects is in the averages of the ratings for the neutral clips. Thus, the dubbing effect has been rated higher on arousal (4.71) in comparison with the voice-over effect (3.9). The statistical tests did not report any significant difference between the values.

The most remarkable and only statistically significant differences are observed in the comparison of the ratings in terms of emotions for the dimension of arousal. According to the results obtained from the self-report instrument, emotions were induced satisfactorily ($p <$ .001).

5.2. Psychophysiological measures

In this section, the results obtained for the measurements of EDA and HR are presented, in graphs and accompanied by a statistical analysis. These results were obtained from the recordings of EDA and HR during the exposure to each stimulus minus the baseline, recorded during the last 3 minutes of the initial relaxation period. In the case of EDA, the results are expressed in microsiemens (μS) and for EDA, in Beats per Minute (bpm).

Due to technical problems during the recording of EDA, the recordings of 9 participants had to be discarded for this measurement, leaving a total of 33. For HR, another 9 participants' recordings were excluded from the analysis. This left the analysis of HR with a total number of 33 participants. These particular files either made no recording (the sensor was misplaced or the connection was lost) or abnormal recordings (EDA is a continuous line that fluctuates, while HR is presented in steps that correspond to the heartbeats). The same participants were not involved in each case, but the number of subjects was the same in each measure, maintaining the external and internal validity of the experiment.

Electrodermal activity. Figure 6 below shows the values for the AST effect. The condition of voice-over obtained higher values (0.5) than that of dubbing (0.35). However, an ANOVA did not show any statistically significant differences between the two AST effects (F (1, 93) = 1.17; p = .28).

Figure 6. EDA for effect. Stimuli average minus 3-minute baseline average

Figure 7 below illustrates the interrelation between the two conditions, emotion and AST effect. The ANOVA did not show any significant differences (F (2, 93) = .38; p = .68). Even a paired comparison with a Bonferroni correction, which was considered to be suitable to assess differences between emotions, did not yield significant differences between AST effects across the conditions.

Figure 7. EDA for AST effect and emotion. Average of stimuli minus 3-minute baseline average

The effect of voice-over resulted in higher values for fear (0.53) and neutral (0.52) emotions compared to the dubbing effect (0.27 and 0.28, correspondingly). In the case of sadness, the average values were very similar for the two AST effects (dubbing = 0.48 and voice-over =

0.45). However, the statistical analysis did not show any significant differences. Again, more experiments mirroring the scientific method are required to test whether a voice-over effect triggers higher emotional activation.

Heart Rate. A difference can be observed between the voice-over effect (0.63) and the dubbing effect (-1.32), as shown in Figure 8 below. However, the results in this condition were not significant ($F (1, 93) = 1.74; p = .19$).

Figure 8. HR for effect. Stimuli averages minus 3-minute baseline average

Figure 9 below shows the interrelation between the two conditions, emotion and AST effect for HR. The ANOVA did not report any significant differences ($F (2, 93) = 1.18; p = .31$). The paired comparison with a Bonferroni correction also did not show any significant differences between AST effects across the three emotions.

Figure 9. HR for AST effect and emotion. Stimuli average minus 3-minute baseline average

It can be observed that the values obtained in the clips treated with a voice-over effect show bigger differences between the stimuli averages and baseline averages for the emotions of sadness (1.18) and fear (1.4) in comparison with the values obtained for the dubbing effect for the same emotions (-2.87 and -1.48, respectively). As in the measurement of EDA, HR did not provide any significant differences between the two effects.

5.3. Correlations

Spearman Rank Correlations were run for each emotion to test the relationship between psychophysiological measures and the ratings of arousal in the questionnaire, and between the two psychophysiological measures (EDA and HR) to see if the reported emotional arousal is

related to the activation of skin conductance and cardiac rhythm. Correlations are presented in the same order.

Little correlation was found between the values obtained through EDA and the ratings obtained in the second scale of the questionnaire, on arousal. The correlations are presented for sadness (r_s (31) = .14, p = .42), and neutral emotion (r_s (31) = .10, p = .57).

Little correlation was found between the values obtained through HR and the ratings obtained in the dimension of arousal. The correlations are presented for sadness (r_s (31) = .01, p = .91), fear (r_s (31) = -.32, p = .06) and neutral emotion (r_s (31) = .03, p = .85).

The relationship between the two psychophysiological measures was assessed for each emotion and little correlation was found. The correlations are presented for sadness (r_s (17) = .43, p = .06), fear (r_s (16) = .20, p = .42) and neutral emotion (r_s (14) = .09, p = .71).

The tests run to assess the relationship between self-report and psychophysiological measures or between the two psychophysiological measures, EDA and HR, did not show any significant correlations for either of the two AST effects, which is a similar outcome to the correlations found in previous research (Fryer, 2013; Ramos, 2015).

6. Conclusion

Audio subtitling is a little-known access service which nonetheless meets many user needs since it is indispensable to access subtitled audiovisual contents for those who do not understand the original language and cannot read the subtitles. Similarly, psychophysiological measures are a little-known methodology for media accessibility despite their potential to provide additional objective measures. This article has presented the results of a study whose innovation lies precisely in these two aspects: the topic and the methodology.

Little is known about user reactions to different audio subtitling strategies. Therefore, the main aim was to research the emotional reactions of persons with sight loss to audio subtitles produced with a dubbing effect as opposed to audio subtitles produced with a voice-

over effect. The main hypothesis was that a dubbing effect would result in a stronger emotional arousal based on the Spanish dubbing tradition to which participants had been exposed. For this purpose, both a self-report questionnaire and psychophysiological measures (EDA and HR) were used. However, the results were different in both methods used and the statistical analysis did not show any significant differences between the two effects. Our hypothesis has not been proven, probably indicating that at this stage both audio subtitling effects may be equally suitable.

The statistical analysis in this type of experiment, carried out with participants with a specific profile, usually has to deal with the issue of sample size. The experiment presented in this article was performed with 42 participants who were blind and partially sighted. Although a bigger sample may indeed may have yielded more powerful statistical results, the number of participants was high if compared to the number of participants in other experiments with blind and partially sighted participants and a methodology based on psychophysiological measures. In Ramos (2015), the measure of HR was carried out with 30 blind and partially sighted participants and in Fryer (2013), the measure of HR and EDA, with 19 blind and partially sighted participants. As a matter of fact, Orero et al. (2018) highlight the challenges of working with participants with special needs, and mention 25 or 30 participants as a good sample of persons with disabilities. It can be argued that the sample in this experiment takes a step further in comparison to previous similar works.

It is worth pointing out that the experiments have demonstrated through the self-report measures that the audio subtitled clips actually induced emotions in blind and partially sighted audiences, as statistically significant differences were observed between emotions. However, the psychophysiological data are highly variable and do not completely correlate with the self-report data or expected results. Many factors, both internal and external to the participant such as concentration or the emotional environment (Ramos 2015), may have had an impact on EDA and HR values. One aspect worth considering for future research is that such

methods seem to have been extensively applied using more immersive stimuli, such as virtual reality for the treatment of phobias (Diemer et al., 2015). It remains to be seen whether the use of regular TV scenes, even if they induce emotions, results in sufficiently strong physiological reactions to provide clear results. In any case, previous research within media accessibility has reported similar results on psychophysiological measures (Fryer, 2013; Ramos, 2015).

The early stage in which the application of such methods to the field of media accessibility and user experience is an invitation for the replication of the experiment as well as the refinement of its parameters. More research with a larger sample, different stimuli, and diverse user profiles is needed in this regard, and the methods and design used in our experiment could be used as inspiration. There is a solid theoretical basis for the study of emotional arousal as an indicator of user experience. Most particularly, the measurement of physiological changes is a promising method that can provide continuous data while the participant is exposed to the stimuli and avoid the subjective nature of questionnaires. To sum up, although the hypothesis has not been confirmed, this research is still valuable from a methodological point of view as it provides the basis for future research and an innovative approach to user testing in audiovisual translation and media accessibility research. This work has been just a first step that calls for more interdisciplinary research with the objective of combining the knowledge of media service designers and the methodological approaches of researchers in psychology, physiology, and neuroscience.

7. References

Backs, R. W., da Silva, S. P., & Han, K. (2005). A comparison of younger and older adults'
        self-assessment manikin ratings of affective pictures. *Experimental Aging Research*, *31*,
        2421–440. doi: 10.1080/03610730500206808

Betella, A., & Verschure, P. F. M. (2016). The Affective Slider: A digital self-assessment
        scale for the measurement of human emotions. *PLoS ONE*, *11*(2), online.
        Doi: 10.1371/journal.pone.0148037

 comparison of younger and older audlts' self-assessment manikin ratings of affective
        pictures. *Experimental Aging Research*, *31*, 2421–440. doi:
        10.1080/03610730500206808

Bradley, M. M., Codispoti, M., Cuthbert, B. N., & Lang, P. J. (2001). Emotion and
        motivation I: Defensive and appetitive reactions in picture processing. *Emotion*, *1*(3),
        276–298. doi: 10.1037/1528-3542.1.3.276

Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: The Self-Assessment Semantic
        Differential Manikin and the semantic differential. *Journal of Behavior Therapy and
        Experimental Psychiatry*, *25*(I), 49–59. doi: 10.1016/0005-7916(94)90063-9

Braun, S., & Orero, P. (2010). Audio description with audio subtitling – an emergent
        modality of audiovisual localisation. *Perspectives: Studies in Translatology*, *18*(3), 173–
        188. doi: 10.1080/0907676X.2010.485687

Cabeza-Cáceres, C. (2013). *Efecte de la velocitat de narració, l'entonació i l'explicitació en
        la comprensió fílmica*. Universitat Autònoma de Barcelona.

Cabeza i Cáceres, C. (2010). Opera audio description at Barcelona's Liceu theatre. In J. Díaz Cintas, A. Matamala, & J. Neves (Eds.), *New Insights into Audiovisual Translation and Media Accessibility* (pp. 227–237). Amsterdam: Rodopi.

Cabeza i Cáceres, C., & Matamala, A. (2008). La audiodescripción de ópera: una nueva propuesta. In A. Pérez-Ugena & R. Vizcaíno-Laorga (Eds.), *Ulises y la comunidad sorda* (pp. 95–108). Observatorio de las Realidades Sociales y de la Comunicación.

Diemer, J., Alpers, G. W., Peperkorn, H. M., Shiban, Y., & Mühlberger, A. (2015). The impact of perception and presence on emotional reactions: A review of research in virtual reality. *Frontiers in Psychology*, *6*, 1–9. Doi: 10.3389/fpsyg.2015.00026

Dillon, C. (2006). *Emotional responses to immersive media*. University of London.

Eardley-Weaver, S. (2014). *Lifting the Curtain on Opera Translation and Accessibility: Translating Opera for Audiences with Varying Sensory Ability*. Durham University.

Fernández-Torné, A., & Matamala, A. (2016). Text to speech vs. human voiced audio descriptions: a reception study in films dubbed into Catalan. *The Journal of Specialised Translation*, (2008), 1–13.

Fresno Cañada, N. (2014). *La (re)construcción de los personajes fílmicos en la audiodescripción. Efectos de la cantidad de información y de su segmentación en el recuerdo de los receptores*. Universitat Autònoma de Barcelona.

Fryer, L. (2013). *Putting It into Words: The Impact of Visual Impairment on Perception, Experience and Presence*. University of London.

Geethanjali, B., Adalarasu, K., Hemapraba, A., Pravin Kumar, S., & Rajasekeran, R. (2017). Emotion analysis using SAM (Self-Assessment Manikin) scale. *Biomedical Resarch, 2017 Special Issue*, S18-S24.

International Organization for Standardization, & International Electrotechnical Commission. (2015). *Information technology - User interface component accessibility - Part 21: Guidance on audio descriptions (ISO/IEC TS 20071-21)*. Retrieved from

https://www.iso.org/obp/ui/fr/#iso:std:iso-iec:ts:20071:-21:ed-1:v1:en

International Organization for Standardization, & International Electrotechnical Commission. (2017). *Information Technology — User interface component accessibility — Part 25: Guidance on the audio presentation of text in videos (including captions, subtitles, and other on screen text) (ISO/IEC 20071-25: 2017)*. Retrieved from https://www.iso.org/standard/69060.html

International Telecommunication Union (Focus Group on Audiovisual Media Accessibility). (2013). *Part 5: Final report of activities: Working Group B "Audio/Video description and spoken captions."* Geneva.

Iturregui-Gallardo, G. (2018). Rendering multilingualism through audio subtitles: shaping a categorisation for aural strategies. *International Journal of Multilingualism*, 1–14. doi: 10.1080/14790718.2018.1523173

Iturregui-Gallardo, G., & Méndez-Ulrich, J. L. (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*. doi: 10.1080/1034912X.2019.1626007

Izard, C. E. (2010). The Many Meanings/Aspects of Emotion: Definitions, Functions, Activation, and Regulation. *Emotion Review*, *2*(4), 363–370. doi: 10.1177/1754073910374661

Jakobson, R. (2000). On linguistic aspects of translation. In L. Venuti (Ed.), *The Translation Studies Reader* (pp. 232–239). London/New York: Routledge. Retrieved from http://www.stanford.edu/~eckert/PDF/jakobson.pdf

Kreibig, S. D. (2010). Autonomic nervous system activity in emotion: A review. *Biological Psychology*, *84*(3), 394–421. doi: 10.1016/j.biopsycho.2010.03.010

Kreibig, S. D., Wilhelm, F. H., Roth, W. T., & Gross, J. J. (2007). Cardiovascular, electrodermal, and respiratory response patterns to fear- and sadness-inducing films.

Psychophysiology, *44*(5), 787.

Kruger, J.-L., Soto-Sanfiel, M. T., Doherty, S., & Ibrahim, R. (2016). Towards a cognitive
audiovisual translatology. In R. Muñoz Martín (Ed.), *Reembedding Translation Process
Research* (pp. 171–194). John Benjamins. doi: 10.1075/btl.128.09kru

Lang, J. P., Bradley, M. M., & Cuthbert, B. N. (2008). *International Affective Picture System
(IAPs): Affective ratings of picture and instruction manual (Technical Report A-8)*.
Gainsville, FL.

Lang, P. J. (1969). *The mechanics of desensitization and the laboratory study of human fear*.
(C. M. Franks, Ed.), *Assessment and status of the behavior therapies*. New York:
McGraw Hill.

Lang, P. J. (1988). What are the data of emotion?. In V. Hamilton, G. H. Bower, & N. H.
Frijda (Eds.), *Cognitive Perspectives on Emotion and Motivation* (pp. 173–191).
Dordrecht: Springer Netherlands. doi: 10.1007/978-94-009-2792-6

Larsen, R. J., & Prizmic-Larsen, Z. (2006). Measuring Emotions: Implications of a
multimethod perspective. In *Handbook of multimethod measurement in psychology.* (pp.
337–351). Washington: American Psychological Association. doi: 10.1037/11383-023

Lindquist, K. A., Kober, H., & Barrett, L. F. (2015). The brain basis of emotion: a meta-
analytic review. *Behavioral and Brain Sciences*, *35*(3), 121–143. doi:
10.1017/S0140525X11000446

Mandryk, R. L., Inkpen, K. M., & Calvert, T. W. (2006). Using psychophysiological
techniques to measure user experience with entertainment technologies. *Behaviour &
Information Technology*, *25*(2), 141–158.

Matamala, A., & Orero, P. (2007). Accessible opera in Catalan: opera for all. In J. Díaz
Cintas, P. Orero, & A. Remael (Eds.), *Media for all: subtitling for the deaf, audio
description, and sign language* (pp. 201–214). Amsterdam: Rodopi.

Matamala, A., Perego, E., & Bottiroli, S. (2017). Dubbing versus subtitling yet again? An

empirical study on user comprehension. *Babel*, *63*(3).

Matamala, A., Soler-Vilageliu, O., Iturregui-Gallardo, G., Jankowska, A., Méndez-Ulrich, J.-
L., Serrano, A. (in press). Electrodermal activity as a measure of emotions in media
accessibility research: methodological considerations. *Journal of Specialised
Translation*.

Miesenberger, K., Klaus, J., & Zagler, W. (2002). Computers helping people with special
needs. In *8th International Conference, ICCHP 2002*. Linz, Austria: Springer.

Montefinese, M., Ambrosini, E., Fairfield, B., & Mammarella, N. (2014). The adaptation of
the Affective Norms for English Words (ANEW) for Italian. *Behavior Research
Methods*, *46*(3), 887–903. doi: 10.3758/s13428-013-0405-3

Nielsen, S., & Bothe, H. H. (2008). Subpal: A device for reading aloud subtitles from
television and cinema. In *CEUR Workshop Proceedings* (Vol. 415). Retrieved from
http://www.scopus.com/inward/record.url?eid=2-s2.0-
84876263872&partnerID=tZOtx3y1

Nowlis, V. (1965). Research with the Mood Adjective Check List. In S. S. Tompkins & C. E.
Izard (Eds.), *Affect, cognition, and personality: Empirical studies* (p. 464). Oxford:
Springer.

O'Hagan, M. (2016). Game Localisation as Emotion Engineering: Methodological
Exploration. In Q. Zhang & M. O'Hagan (Eds.), *Conflict and Communication: A
Changing Asia in a Globalising World* (pp. 123–144). New York: Nova.

Orero, P. (2007). Audiosubtitling: A possible solution for opera accessibility in Catalonia.
Tradterm, *13*, 135. doi: 10.11606/issn.2317-9511.tradterm.2007.47470

Orero, P. (2011). The audiodescription of spoken, tactile and written language in *Be with me*.
In A. Matamala, J.-M. Lavaur, & A. Serban (Eds.), *Audiovisual Translation in Close-up*
(pp. 239–255). Bern: Peter Lang.

Orero, P., & Matamala, A. (2007). Accessible opera: Overcoming linguistic and sensorial

barriers. *Perspectives: Studies in Translatology*, *15*(4), 262–277.

Orero, P., Doherty, S., Kruger, J.-L., Matamala, A., Pedersen, J., Perego, E., … Szarkowska, A. (2018). Conducting experimental research in audiovisual translation (AVT): A position paper. *JoSTrans*, *30*, 105–126.

Perego, E., Del Missier, F., & Bottiroli, S. (2014). Dubbing versus subtitling in young and older adults: cognitive and evaluative aspects. *Perspectives*, *6623*(July), 1–21. doi: 10.1080/0907676X.2014.912343

Perego, E., Del Missier, F., Porta, M., & Mosconi, M. (2010). The cognitive effectiveness of subtitle processing. *Media Psychology*, *13*(3), 243–272. doi: 10.1080/15213269.2010.502873

Pishghadam, R., Firooziyan, A., & Esfahani, P. (2017). Introducing emotioncy as an effective tool for the acceptance of persian neologisms. *Language Related Research.*, *8*(5).

Plutchik, R. (2009). Emotions: A General Psychoevolutionary Theory. In K. R. Scherer & P. Ekman (Eds.), *Approaches to emotion*. New York, London: Psychology Press.

Ramos Caro, M. (2016). *La traducción de los sentidos*. Munich: LINCOM GmbH.

Ramos, M. (2015). The emotional experience of films: Does audio description make a difference?. *Translator*, *21*(1), 68–94. doi: 10.1080/13556509.2014.994853

Remael, A., Reviers, N., & Vercauteren, G. (2014). *ADLAB Audio Description guidelines*. European Union under the Lifelong Learning Programme (LLP).

Reviers, N., & Remael, A. (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.

Romero-Fresco, P., & Fryer, L. (2014). Could audio described films benefit from audio introductions? An audience response study. *Journal of Visual Impairment & Blindness*, *107*(4), 287–295.

Rooney, B., Benson, C., & Hennessy, E. (2012). The apparent reality of movies and

emotional arousal: A study using physiological and self-report measures. *Poetics*, *40*(5), 405–422. doi: 10.1016/j.poetic.2012.07.004

Rottenberg, J. (2003). When emotion goes wrong: Realizing the promise of affective science. *Clinical Psychology: Science and Practice*, *10*(2), 227–232. doi: 10.1093/clipsy/bpg012

Rottenberg, J., Ray, R. D., & Gross, J. J. (2007). Emotion Elicitation Using Films. In J. A. Coan & J. J. B. Allen (Eds.), *Handbook of Emotion Elicitation and Assessment* (pp. 9–28). New York: Oxford University Press.

Scherer, K. R. (2005). What are emotions? And how can they be measured?. *Social Science Information*, *44*(4), 695–729. doi: 10.1177/0539018405058216

Sepielak, K. (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.

Stearns, P. N., & Stearns, C. Z. (1985). Emotionology: Clarifying the history of emotions and emotional standards. *The American Historical Review*, *90*(4), 813–836.

Stemmler, G. (2004). Physiological Processes during Emotion. In P. Philippot, & R. S. Feldman (Eds.), *The Regulation of Emotion* (pp. 33–70). Mahwah, NJ and London, UK: Lawrence Erlbaum Associates. Doi: 10.4324/9781410610898

Szarkowska, A. (2011). Text-to-speech audio description: Towards wider availability of AD. *The Journal of Specialised Translation*, *15*, 142–162. Retrieved from http://www.jostrans.org/issue15/art_szarkowska.pdf

Szarkowska, A., & Mączyńska, M. (2011). *Text-to-speech Audio Description with Dudio Subtitling to a Non-Fiction Film . A Case Study Based on* La Soufriere *by Werner Herzog*. University of Warsaw.

Udo, J. P., Acevedo, B., & Fels, D. I. (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. doi:

10.1177/0264619609359753

Udo, J. P., & Fels, D. (2009). Re-fashioning fashion: An exploratory study of a live audio

described fashion show. *Universal Access in the Information Society*, *9*(1), 63–75. doi:

10.1007/s10209-009-0156-1

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*(February). doi: 10.3389/fpsyg.2016.00180

Watson, D., & Clark, L. A. (1999). *The PANAS-X: Manual for the Positive and Negative*

Affect Schedule - Expanded Form. Iowa City: Department of Psychological & Brain

Sciences Publications.

Westermann, R., Spies, K., Stahl, G., & Hesse, F. W. (1996). Relative effectiveness and

validity of mood induction procedures: A meta-analysis. *European Journal of Social*

Psychology, *26*(4), 557–580. doi: 10.1002/(SICI)1099-0992(199607)26:4<557::AID-

EJSP769>3.0.CO;2-4

Wilson, G. M., & Sasse, M. A. (2006). Do users always know what's good for them?

Utilising physiological responses to assess media quality. In S. McDonald, Y. Waern, &

G. Cocktrn (Eds.), *People and Computers XIV — Usability or Else! Poceedings of HCI*

2000 (pp. 327–339). London: Springer London.

Clip	Emotion	Effect
1	Sadness	VO
2	Sadness	Dubbing
3	Fear	VO
4	Fear	Dubbing
5	Neutral	VO
6	Neutral	Dubbing

Table 1. Stimuli of the experiments