

# Audio subtitling: voicing strategies and their effect on film enjoyment

GONZALO ITURREGUI-GALLARDO<sup>1</sup>, ANNA SERRANO<sup>2</sup>, JORGE LUIS MÉNDEZ-ULRICH<sup>3</sup>, ANNA JANKOWSKA<sup>4</sup> AND OLGA SOLER-VILAGELIU<sup>5</sup>

<sup>1</sup> Universitat Autònoma de Barcelona, Bellaterra, Spain  
e-mail: gonzalo.iturregui@uab.cat

<sup>2</sup> Universitat Autònoma de Barcelona, Bellaterra, Spain  
e-mail: anna.serranora@e-campus.uab.cat

<sup>3</sup> Universitat Autònoma de Barcelona, Bellaterra, Spain  
e-mail: jorgeluis.mendez@uab.cat

<sup>4</sup> Universitat Autònoma de Barcelona, Bellaterra, Spain  
e-mail: anna.jankowska@uab.cat

<sup>5</sup> Universitat Autònoma de Barcelona, Bellaterra, Spain  
e-mail: olga.soler@uab.cat

## Abstract

Media Accessibility, understood as a branch of Audiovisual Translation studies devoted to the study of access services to media, has experienced a growth in research in recent years. Even if some of the techniques in this field have produced a great number of works, others are still underresearched. This article reports on a pilot study carried out within the NEA Project<sup>1</sup>. The aim of the study is to compare different AST delivery styles in an experiment during which data will be gathered through subjective and objective measures. This article reports on the results of a pilot study in which the research procedure was set up and tested. Our main findings show that both subjective and objective assessment methods are valid to evaluate emotional experience induced by films.

<sup>1</sup> Funded by the Spanish Ministry of Economy, Industry and Competitiveness (FFI2015-64038-P, MINECO/FEDER, UE) within the research group Transmedia Catalonia (2014SGR0027)

## 1 Introduction

Access services have received a lot of attention in Audiovisual Translation research recently (Maszerowska, Matamala, & Orero, 2014; Matamala & Orero, 2016; Romero-Fresco, 2015), developing a new branch known as Media Accessibility (see Greco, 2016). However, there are still some modalities such as audio subtitling (AST), which have attracted less attention from researchers.

Audio subtitling is understood as a hybrid technique that drags characteristics from other AVT strategies such as dubbing, voice-over or subtitling (Braun & Orero, 2010). It is based on the “aurally rendered and recorded version of subtitles with a film” (Reviere & Remael, 2015: 52). Whereas subtitling countries that have found in AST a tool to provide access for the visually impaired (see Hanzliček, Matoušek, & Tihelka, 2008; Ljunglöf, Derbring, & Olsson, 2012; Nielsen & Bothe, 2008; Theunisz, 2002), in dubbing countries its use is proves

helpful in multilingual contents. Multilingualism is often maintained in films by using subtitles for secondary languages, which have to be adapted for people with special needs, such as the visually impaired and people with reading problems (Ljunglöf et al., 2012).

For the purposes of this study we consider two different strategies of AST vocal delivery, which have been mentioned in previous works (Braun & Orero, 2010; ISO & IEC, 2016; Remael, 2014). The AST track can be delivered either with a dubbing effect or a voice-over effect. In the former the AST is delivered by imitating orality and the original can no longer be heard. On the other hand, in the latter, the AST shows less prosodic features and the original can be heard.

The paper reports on the results of an experiment carried out as a pilot study in order to test the experimental procedure. Below we present the NEA project, previous research in the field of AST,

methodology, experimental design and preliminary results.

## 2 NEA PROJECT

NEA is a project that is set to shed light on new approaches to accessibility by looking into its various aspects. One of them is the study of vocal delivery in AST, namely the dubbing and the voice-over effect, described in this article.

We argue that the two different strategies used for voicing AST in multilingual content will have an impact on the audience. User experience (UX) (see Albert & Tullis, 2013) is the approach taken to test the effect caused in the spectator. By conceiving the filmic experience as a way of entertainment which aims at the engrossment of the spectator (Vorderer & Hartmann, 2008), the experiment aims at the study of presence and emotional arousal (see Lombard & Jones, 2015). It will combine more traditional self-report instruments and psychophysiological measures (skin conductance and heart rate), which move to a more objective approach of such issue. Within the experiment the impact of the two effects will be put to the test with blind and partially sighted and, also, compared to the impact of regular written subtitles on sighted audiences.

## 3 Related Work

It could be argued that the study on AST counts with still a modest amount of works. However, previous research has dealt with different aspects and applications of AST. Some works have centred on systems that allow for the automatic reading of AST for regular TV broadcasting, particularly in subtitling countries such as the Netherlands (Theunisz, 2002) or Denmark (Thrane, 2013). Such systems are based on text-to-speech (TTS) and optical character recognition (OCR) technologies to cater for more accessibility and autonomy for users with specific needs.

It is important, as well, to mention previous experiences in genres different from television, such as opera (Orero, 2007). This study presented a case study in which audio surtitles could overcome the language barrier faced by people with visual impairments.

In the field of cinema, AST have been studied in combination with AD, particularly required in multilingual films. Of special relevance is the information provided in proposals for making accessible text on screen in films (Matamala, 2014) and the analyses of commercialised multilingual films with AD and AST (Benecke, 2012; Braun & Orero, 2010; Harrouet, 2016; Remael, 2012). AST have been also

studied in comparison to other audiovisual translation modes, such as voice-over and subtitling, adapted for AD (Jankowska, Szarkowska, & Mentel, 2015).

The conclusions in the previous works highlight the lack of compromise in the employment of audio subtitling. Indeed, even if few scholars have worked on some guidelines (see, for instance, Remael, 2014), their delivery and presentation remains unclear when considering the contents of standards and guidelines (Iturregui-Gallardo, forthcoming). Although it is true that there is a division between the two so-called effects (Braun & Orero, 2010; ISO & IEC, 2016; Remael, 2014), the definition of the strategies and their impact on the experience of the audience have not been analysed in deep yet.

When it comes to incorporating psychophysiology in the study of the impact of audiovisual contents, some previous research was carried out; for example in videogames (O'Hagan, 2016) and film scenes (Rooney, Benson, & Hennessy, 2012). Of special relevance, though, are those focused on the effect of such stimuli on audiences with visual impairments. The experimental design of the pilot study presented in this paper is based on two studies presenting audiovisual contents with AD. Fryer (2013) focused on the way participants felt immersed in the audiovisual content by means of analysing the levels of presence (Lombard & Jones, 2015). Ramos (2015), in turn, collected data for emotional arousal provoked by scenes portraying different emotions.

## 4 Methodology

It is important first of all to recall on the fact that this experiment is based on two different kinds of methods to study user experience: questionnaires and psychophysiological measurements.

Three questionnaires were presented to the participants along the entire experiment. Before the experiment begins, the Toronto Alexithymia Scale (TAS-20) is presented. It assesses the possible alexithymia of the subject, which is the incapacity to identify their own emotions and to express them (see Taylor et al., 1988). After each of the audiovisual stimuli the following questionnaires were presented. The Self Assessment Mannequin (SAM) questionnaire, which evaluates emotion in three factors: pleasure, arousal and dominance. It consists of three graphic scales with representative icons that represent different levels of emotional affect (see Lang & Vaitl, 1988); and the Narrative Engagement Scale: it measures the subjective engagement perception. It contains 12 items that relate to the dimensions of Narrative Understanding,

Attentional Focus, Narrative Presence and Emotional Engagement. The questionnaire is answered by providing the options “yes” or “no”.

For the psychophysiological measurements two different devices were used: the Captiv L-7000, both the software and the T-sense sensors for galvanic skin response (GSR) and heart rate (HR); and the Tobii eye-tracker. Since the ultimate participants will be visually impaired people eye-tracking measuring was excluded. However, its software was used as the experimental procedure which presented film fragments and instructions for the subject.

### 4.1 Experimental Protocol

The protocol followed in the pilot study, which is based on the visualisation of 3 film fragments, had a total duration of 30 minutes.

Firstly, the participant arrives and a concise explanation of the experiment is provided and the participant responds to the TAS-20 test. Afterwards, the participant is invited to a brief relaxation period, induced by mood music, which lasts eight minutes. The relaxation period exists in order to unify the participants’ emotional state and to record their base line. After the relaxation, the first film fragment is projected. Then, the participant responds to the SAM and to the engagement questionnaire marked with an A and submits to a four-minute relaxation. The protocol is repeated for fragments two and three, marked as B and C respectively. During the presentation of each video, the GSR, and the HR, were recorded using the Captiv Software V1.5. Descriptive analyses of the SAM subscales means and integration analyses of the GSR and HR during the videos were carried out using the IBM SPSS v.22 for Windows.

This pilot experiment was tested on five participants between 20 and 27 years old. The instructions of the experiment, as well as the questionnaires were all presented in Spanish.

### 5 Experimental Results

First, the analysis of the emotional assessment obtained through the SAM scale suggest that the emotion generated by each video used in the experiment protocol was different and consistent with that which was intended to induce, as reflected by the average scores in emotional arousal, valence and dominance, which are summarized in Table 1.

	Arousal	Valence	Dominance
Happiness	5.0	6,40	6.20
Anger	6.60	2.20	4.40
Sadness	5.80	2,4	5.0

Table 1. SAM subscales mean for each video.

Regarding the psychophysiological assessment of the arousal induced by the videos, as shown in Table 2 and Table 3, our preliminary results suggest the existence of different patterns of psychophysiological reactivity based on the emotional content of each video, given the differences observed in the analysis of the integration of the GSR and the HR, even the small sample size (n=5) does not allow to assess the statistical significance of these findings.



Table 2. Integration analysis of HR during each video (arbitrary units).

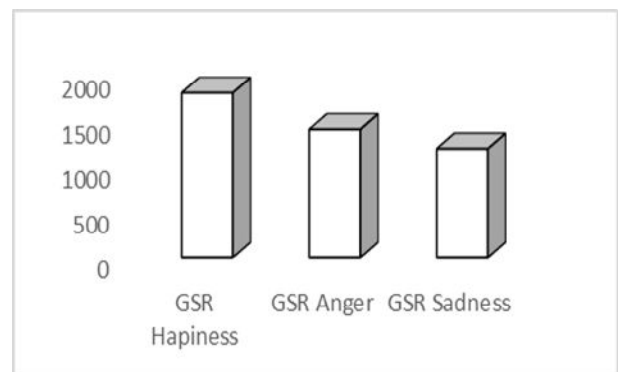


Table 3. Integration analysis of GSR during each video (arbitrary units).

### 6 Conclusion

In conclusion, despite the methodological limitations of our work (especially concerning the sample size), our results suggest that both the subjective emotional assessment performed by the SAM scale, as well as the objective physiological measures are a valid method to evaluate the emotional experience induced by different

video fragments. This justifies further research to improve the accessibility of films for blind and visually impaired people through audio subtitles, using adapted versions of the SAM scale, in what our team is working.

The promising outcome of this pilot study has served as a tool to validate the experimental protocol for the ultimate experiment with people with visual impairments. It is worth reminding the novelty and the quasi inexistence of previous experiences where such measurements were applied in the media accessibility field. However, we encourage the incorporation of such methods in order to provide a different approach to the way users interact with such contents, idiosyncratic to nowadays society.

## References

- Albert, W., & Tullis, T. (2013). *Measuring the User Experience : Collecting, Analyzing, and Presenting Usability Metrics* (2nd ed.). Burlington, MA: Morgan Kaufmann.
- Benecke, B. (2012). Audio description and audio subtitling in a dubbing country : Case studies. In E. Perego (Ed.), *Emerging topics in translation: Audio description* (pp. 99–104). Trieste: EUT Edizioni Università di Trieste.
- Braun, S., & Orero, P. (2010). Audio description with audio subtitling – an emergent modality of audiovisual localisation. *Perspectives: Studies in Translatology*, 18(3), 173–188. <https://doi.org/10.1080/0907676X.2010.485687>
- Fryer, L. (2013). *Putting It Into Words : The Impact of Visual Impairment on Perception , Experience and Presence*. University of London.
- Greco, G. M. (2016). On Accessibility as a Human Right, with an Applicatin to Media Accessibility. In A. Matamala & P. Orero (Eds.), *Researching Audio Description: New Approaches* (pp. 11–34). London: Palgrave Macmillan.
- Hanzliček, Z., Matoušek, J., & Tihelka, D. (2008). Towards automatic audio track generation for Czech TV broadcasting: Initial experiments with subtitles-to-speech synthesis. *International Conference on Signal Processing Proceedings, ICSP*, 2721–2724. <https://doi.org/10.1109/ICOSP.2008.4697710>
- Harrouet, C. (2016). *Multilingualism in Audio-Description: Audio-Subtitling. MA Thesis*. Roehampton University. <https://doi.org/10.1002/ejoc.201200111>
- ISO, & IEC. (2016). ISO/IEC 20071-25. 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).
- Iturregui-Gallardo, G. (forthcoming). *Audio subtitling: strategies for voicing subtitles and their effect on film enjoyment and comprehension*. Autonomous University of Barcelona.
- Jankowska, A., Szarkowska, A., & Mentel, M. (2015). Why big fish isn't a fat cat? Adapting voice-over and subtitles for audio description in foreign films. In L. Bogucki & M. Deckert (Eds.), *Accessing Audiovisual Translation, Series: Lodz Studies in Language, vol. 35*. Frankfurt am Main, Berlin, Bern, Bruxelles, New York, Oxford, Wien: Peter Lang. <https://doi.org/10.3726/978-3-653-04906-0>
- Lang, P. J., & Vaitl, D. (1988). The international affective picture system [Photografic slides]. Gainesville, FL: University of Florida, Centre for Research in Psychophysiology.
- Ljunglöf, P., Derbring, S., & Olsson, M. (2012). A Free and Open-Source Tool That Reads Movie Subtitles Aloud. In *Workshop on Speech and Language Processing for Assistive Technologies* (pp. 1–4).
- Lombard, M., & Jones, M. T. (2015). Defining Presence. In F. Biocca, M. Lombard, J. Freeman, W. A. Ijsselstein, & R. J. Schaevitz (Eds.), *Immersed in Media: Telepresence Theory, Measurement & Technology* (pp. 13–34). Cham: Springer International Publishing.
- Maszerowska, A., Matamala, A., & Orero, P. (2014). *Audio Description: New perspectives illustrated*. Amsterdam: John Benjamins Publishing Company.
- Matamala, A. (2014). Audio Describing Text on Screen. In A. Maszerowska, A. Matamala, & P. Orero (Eds.), *Audio Description: New Perspectives Illustrated* (pp. 103–120). Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Matamala, A., & Orero, P. (2016). *Researching Audio Description*. London: Palgrave Macmillan.
- 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>
- 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. 1–15). New York: Nova.
- Orero, P. (2007). Audiosubtitling: a possible solution for opera accessibility in Catalonia. *Tradterm*, 13, 135. <https://doi.org/10.11606/issn.2317-9511.tradterm.2007.47470>
- Ramos, M. (2015). The emotional experience of films: Does audio description make a difference?

- Translator*, 21(1), 68–94.  
<https://doi.org/10.1080/13556509.2014.994853>
- Remael, A. (2012). Audio Description with Audio Subtitling for Dutch Multilingual Films: Manipulating Textual Cohesion on Different Levels. *Meta: Journal Des Traducteurs*, 57(2), 385–407. <https://doi.org/10.7202/1013952ar>
- Remael, A. (2014). Combining audio description with audio subtitling. In A. Remael, N. Reviere, & G. Vercauteren (Eds.), *ADLAB Audio Description Guidelines*. Retrieved from <http://www.adlabproject.eu/Docs/adlabbook/index.html#combining-ad>
- Reviere, 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. (2015). *The Reception of Subtitles for the Deaf and Hard of Hearing in Europe*. (P. Romero-Fresco, Ed.). Bern, Berlin, Bruxelles, Frankfurt am Main, New York, Oxford, Wien: Peter Lang.
- 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.  
<https://doi.org/10.1016/j.poetic.2012.07.004>
- Taylor, G. J., Bagby, R. M., Ryan, D. P., Parker, J. D., Doody, K. F., & Keefe, P. (1988). Criterion validity of the Toronto Alexitymia Scale. *Psychosomatic Medicine*, 50(5), 500–9. Retrieved from <https://xpv.uab.cat/DanaInfo=.aIoukmmzzHx5tpMq32,SSL+pubmed?pmid=3186893#>
- Theunisz, M. (2002). Audio subtitling : a new service in the Netherlands making subtitled programmes accessible. Retrieved from [www.sb-belang.nl](http://www.sb-belang.nl)
- Thrane, L. K. (2013). *Text-to-Speech on Digital TV: An Explanatory Study of Spoken Subtitles on DRISyn*. IT University of Copenhagen.
- Vorderer, P., & Hartmann, T. (2008). Entertainment and enjoyment as media effects. In J. Bryant & M. B. Oliver (Eds.), *Media effects: Advances in theory and research* (pp. 532–550). London: Routledge.