

First steps of the NEA project: audio subtitling

G. Iturregui-Gallardo, A. Matamala, O. Soler-Vilageliu

Universitat Autònoma de Barcelona, TransMedia Catalonia

gonzalo.iturregui/anna.matamala/olga.soler@uab.cat

Languages and the Media, 2-4 November 2016

FFI2015-64038-P (MINECO/FEDER, UE), 2014SGR0027, 2016FI_B 00012

Outline

- The NEA (Nuevos Enfoques sobre la Accesibilidad/New Approaches to Accessibility) project
 - Why this project?
 - Aims
 - Psychophysiological measures
- First steps: audio subtitling (PhD)
 - Literature review
 - Corpus analysis: preliminary results

NEA: why this project?

Underresearched aspects of audio description

- Audio component
- Audio subtitling
- New methodologies to assess emotional arousal

Aims

- Descriptive analyses with a focus on audio subtitling
- Analyse the impact of audio on user experience
- Apply new psychophysiological measures to assess it

➤ And transfer all this to research community, teaching practices and society

Psychophysiological measures

- Assessment of reception:
 - Going beyond comprehension and preferences into emotional arousal/immersion/presence >
- Psychophysiological measures:
 - Electrodermal activity and heart rate
 - Facial expression (web camera)

Captiv L7000



Previous work

- Psychophysiological measures
 - Kurz, I., 2003 (Interpreting)
 - Baumgartner, Esslen & Jäncke, 2006 (Music and pictures)
 - Sohn, Sokhadze & Watanuki, 2001 (Children)
 - Daltrozzo et al., 2010 (Patients in comma)
- Applied to audiovisual materials
 - O'Hagan, 2016 (Videogames)
 - Brumbaugh et al., 2013 (Emotional video scenes)
 - Rooney, Benson & Hannessy (2D and 3D films)
 - Codispoti, Surcinelly & Baldaro, 2008 (Film genres)

Previous work in AD

- AD on HR
 - The emotional experience of films: Does audio description make a difference? Ramos, 2015.
- AD on EDA and HR
 - Putting It Into Words: The Impact of Visual Impairment on Perception, Experience and Presence. Fryer, 2013.

First step: audio subtitling

- PhD on audio subtitling by Gonzalo Iturregui-Gallardo
- **Aim:**
 - Compare reception of audio subtitling depending on voicing strategy: dubbing vs. voice-over effect
- **Focus:**
 - Users' emotional arousal by using psychophysiological measures

Current work

- Since February 2016
 - Project definition: Audio subtitling when combined with AD
 - Literature review
 - Automatic AST
 - TTS vs. human voices
 - Dubbing vs. voice-over effect
 - Preliminary corpus analysis

Stage 1: literature review

- Limited number of works on AST, and often linked to AD
- Aspects discussed in the literature:
 - **Automatic AST**
 - Thrane, 2013; Derbring, Ljunglöf & Olsson, 2009; Mihkla et al., 2014; Hanzliček, Matoušek & Tihelka, 2008
 - **TTS vs. human-voiced AST**
 - Szarkowska & Jankowska, 2012
 - **Dubbing vs. voice-over effect**
 - Braun & Orero, 2010
- AST in
 - **Regulations and standards**
 - **Guidelines**

Stage 1: regulations

30 countries¹ (Europe and English-speaking countries overseas)

Regulations

- Bulgaria:

Electronic Communications Act, 2007

- Flemish Community:

Decree of 27th March 2009 on the radio and television

- Poland:

Act on radio and television, 1992 (amendment)

- Finland:

Act on Television and Radio Operations, 2010

- Sweden:

Radio and Television Act, 2010; The Broadcasting Authority obligations, 2011; Proposed Government Bill, 2014-2019

Standards

- AENOR, Spain

- Ofcom, UK

- ITC Guidance, UK

- ISO/IEC 20071-25: “Information Technology – User interface component accessibility”

- Automatic and human-voiced AST

- Combination with AD

- Pre-recorded and live AST

- Use of voice and prosody

- Important: call for more research

¹Including the region of Catalonia, Spain and those belonging to the European Union.

Stage 1: guidelines

Guidelines and best practices containing information on on-screen text

- BAI Guidelines, Ireland
- ADC Guidelines, USA
- AFB Guidelines, USA
- ADB Guidelines, Australia
- Audio description guidelines for the Greek, Greece
- MAC Descriptive Video Production and Presentation of Best Practices Guide for Digital Environments, Canada
- Described Video Best Practices, Canada
- Bayerischer Rundfunk Guidelines, Germany

Guidelines with further information on AST

- ACB's Audio Description Project, USA
- Netflix, USA
- ADLAB Project, Europe

Stage 2: corpus analysis

- Scarce literature. Preliminary small-scale study using a qualitative approach
- Inspired by previous research
 - Benecke, 2012; Remael, 2012; Orero and Braun, 2010; Matamala, 2014
- Preliminary corpus (to be expanded):
 - Multilingual films
 - 21st century

Stage 2: corpus analysis

- **Materials:**

- *Everything is Illuminated* (L. Schreib, 2005) (English, Ukranian, Russian)
- *Syriana* (S. Caghan, 2005) (English, Arabic, Farsi, Urdu, Mandarin, French)
- *The Science of Sleep* (M. Gondry, 2006) (English, French, Spanish)
- *Borat* (L. Charles, 2006) (English, Hebrew, Armenian, Romanian, Polish)

- **Next:**

- *Munich* (S. Spielberg, 2005) (English, German, French, Hebrew, Arabic, Italian, Greek, Russian)
- *My name is Khan* (K. Johar, 2010) (Hindi, Urdu, English)
- *Slumdog Millionaire* (D. Boyle & L. Tandan, 2008) (English, Hindi, French)

Stage 2: matrix of analysis

Title (Year)		
Director		
Run time		min
Country		
Main language		
Other languages		
Scenes with AST		
AST time		min
Character recognition	Characters (Sex & Age)	
	Voices (Sex & Age)	
Assignment	Reported / Direct speech	
	Combination with AD	
Audibility	Original audibility	
	Dubbing / Voice-over effect	
Prosody (performance)		
Isochrony (delay)		
Reading of written subtitles		
Omission / Expansion		
Observations:		

Stage 2: Initial results

Character recognition

- Number of characters and voice talents
- Trends:
 - Sex of the AD voice and the AST voices
 - Character – voice. Budget constraints
 - Voice talent has more than one character
 - Scenes with characters of the same sex are more complex

Stage 2: Initial results

Audibility

- Volumes and combination of the original track and the AD/AST track
- Trends:
 - Merging vs. overlapping
 - Voice-over effect
 - Surprise effect

Stage 2: initial results

Assignment

- Audio describer's introduction of AST and combination with AD content
- Trends:
 - Direct speech
 - When scene could be confusing, AD adds information on the conversational situation
 - “she/he says” or “X says”

Stage 2: initial results

Prosody

- The use of the voice. Differences in prosodic features which take distance from neutrality
- Trends:
 - Voices are performed
 - Use of different prosodic characteristics for different characters
 - According to film genre and tone
 - AST as a part of the artistic experience (beyond an accessibility service)

Stage 2: initial results

Isochrony

- Synchronisation between the length of the original dialogue and the overlapping AST
- Trends:
 - AST comes before the original line
 - No perfect isochrony
 - Depends on time constraints and complexity of the scene
 - Isochrony and character recognition
 - Lack of delay and bad audibility of the original line

Future stage: experiments

- Design
 - Captiv L7000
 - Definition of variables
- Selection of materials
- Preparation of materials
- Contact participants

References



- Bailey, J., Bailenson, J., Won, A., Flora, J., & Armel, K. (2011). Presence and Memory: Immersive Virtual Reality Effects on Cued Recall. Proceedings of the International Society for Presence Research Annual Conference.
- Baumgartner, T., Esslen, M., & Jäncke, L. (2006). From emotion perception to emotion experience: Emotions evoked by pictures and classical music. *International Journal of Psychophysiology*, 60(1), 34–43. <http://doi.org/10.1016/j.ijpsycho.2005.04.007>
- 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.
- Biocca, F., & Kim, T. (1997). Telepresence via television: Two dimensions of telepresence may have different connections to memory and persuasion. *Journal of Computer-Mediated Communication*, 3(2).
- Braun, S., & Orero, P. (2010). Audio description with audio subtitling – an emergent modality of audiovisual localisation. *Perspectives: Studies in Translatology*, 18(3), 173–188. <http://doi.org/10.1080/0907676X.2010.485687>
- Brumbaugh, C. C., Kothuri, R., Marci, C., Siefert, C., & Pfaff, D. D. (2013). Physiological correlates of the big 5: Autonomic responses to video presentations. *Applied Psychophysiology Biofeedback*, 38(4), 293–301. <http://doi.org/10.1007/s10484-013-9234-5>
- Busselle, R., & Bilandzic, H. (2009). Measuring Narrative Engagement. *Media Psychology*, 12(4), 321–347. <http://doi.org/10.1080/15213260903287259>
- Cho, D., Park, J., Kim, G., Hong, S., Han, S., & Lee, S. (2003). Dichotomy of presence elements: The where and what. In *Proceedings of the IEEE Virtual Reality* (pp. 273–274).
- Codispoti, M., Surcinelli, P., & Baldaro, B. (2008). Watching emotional movies: Affective reactions and gender differences. *International Journal of Psychophysiology*, 69(2), 90–95. <http://doi.org/10.1016/j.ijpsycho.2008.03.004>
- Daltrozzo, J., Wioland, N., Mutschler, V., Lutun, P., Calon, B., Meyer, A., ... Kotchoubey, B. (2010). Emotional electrodermal response in coma and other low-responsive patients. *Neuroscience Letters*, 475(1), 44–47. <http://doi.org/10.1016/j.neulet.2010.03.043>
- Derbring, S., Ljunglöf, P., & Olsson, M. (2009). SubTTS: Light-Weight Automatic Reading of Subtitles. *Nordic Conference on Computational Linguistics*, 272–274.
- Fryer, L. (2013). *Putting It Into Words : The Impact of Visual Impairment on Perception , Experience and Presence*. University of London.
- Fryer, L., & Freeman, J. (2013). Visual impairment and presence. In H. Witchel (Ed.), *Proceedings of the 2013 Inputs-Outputs Conference on An Interdisciplinary Conference on Engagement in HCI and Performance - Inputs-Outputs '13* (pp. 1–5). <http://doi.org/10.1145/2557595.2557599>
- Hanzlíček, 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. <http://doi.org/10.1109/ICOSP.2008.4697710>
- Khalfa, S., Isabelle, P., Jean-Pierre, B., & Manon, R. (2002). Event-related skin conductance responses to musical emotions in humans. *Neuroscience Letters*, 328(2), 145–149. [http://doi.org/10.1016/S0304-3940\(02\)00462-7](http://doi.org/10.1016/S0304-3940(02)00462-7)

References



- Kurz, I. (2003). Psychophysiological stress during simultaneous interpreting: a comparison of experts and novices. *The Interpreters' Newsletter*, 12, 51–67.
- Lessiter, J., Freeman, J., Keogh, E., & Davidoff, J. (2001). A Cross-Media Presence Questionnaire: The ITC-Sense of Presence Inventory. *Presence: Teleoperators and Virtual Environments*, 10(3), 282–297. <http://doi.org/10.1162/105474601300343612>
- 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.
- Mihkla, M., Hein, I., Kiissel, I., R app, A., Sirts, R., & Valdna, T. (2014). A System of Spoken Subtitles for Estonian Television. *Human Language Technologies - the Baltic Perspective*, 268, 19–26. <http://doi.org/10.3233/978-1-61499-442-8-19>
- 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.
- Picard, R. W. (2000). Toward computers that recognize and respond to user emotion. *IBM Systems Journal*, 39(3 & 4), 705–719. <http://doi.org/10.1147/sj.393.0705>
- Price, M. (2006). *The Relation of Presence and Virtual Reality Exposure for Treatment of Flying Phobia*. Georgia State University.
- Ramos, M. (2015). The emotional experience of films: Does audio description make a difference? *Translator*, 21(1), 68–94. <http://doi.org/10.1080/13556509.2014.994853>
- Ramos Caro, M. (2016). Testing audio narration: the emotional impact of language in audio description. *Perspectives*, 24(4), 1–29. <http://doi.org/10.1080/0907676X.2015.1120760>
- 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. <http://doi.org/10.7202/1013952ar>
- 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. <http://doi.org/10.1016/j.poetic.2012.07.004>
- Slater, M., Usoh, M., & Steed, A. (1994). Depth of Presence in Immersive Virtual Environments, *Presence: Teleoperators and Virtual Environments*. MIT Press, 3(2), 130–144.
- Sohn, J. H., Sokhadze, E., & Watanuki, S. (2001). Electrodermal and cardiovascular manifestations of emotions in children. *Journal of Physiological Anthropology and Applied Human Science*, 20(2), 55–64. <http://doi.org/10.2114/jpa.20.55>
- Szarkowska, A., & Jankowska, A. (2012). Text-to-speech Audio Description of Voiced-over Films. A Case Study of Audio Described *Volver* in Polish. In E. Perego (Ed.), *Emerging Topics in Translation: Audio description* (pp. 81–98). Trieste: EUT Edizioni Università di Trieste.
- Thrane, L. K. (2013). *Text-to-Speech on Digital TV: An Explanatory Study of Spoken Subtitles on DR1Syn*. IT University of Copenhagen.

First steps of the NEA project: audio subtitling

G. Iturregui-Gallardo, A. Matamala, O. Soler-Vilageliu

Universitat Autònoma de Barcelona, TransMedia Catalonia

gonzalo.iturregui/anna.matamala/olga.soler@uab.cat

Languages and the Media, 2-4 November 2016

FFI2015-64038-P (MINECO/FEDER, UE), 2014SGR0027, 2016FI_B 00012