

# Speech and translation technologies for voice-over and audio description: final results of the ALST project

**Anna Matamala (Universitat Autònoma de Barcelona)**

**[anna.matamala@uab.cat](mailto:anna.matamala@uab.cat)**

**6th International Conference Media for All  
Western Sydney University, 17-18/09/2015.**

**Funding: FFI-2012-31024, 2014SGR0027**



# The ALST project: aims

- Speech recognition
- Machine translation
- Speech synthesis

in audio description

in voice-over

# The ALST project: why

- No development but implementation of technologies
- Two oral transfer modes (linguistic/sensorial accessibility)
- Increase quantity? Comparable quality?
- Impact on describers and on audiences
- Limitations: three years, very limited funding
- Project team and 2 PhD students

# Speech recognition in voice-over

- 10 professional transcribers in three scenarios:
  - Manual transcription
  - Respeaking
  - Revision of automatic speech recognition
- Quantitative data: time ratio, output quality (NER)
- Qualitative data: opinions on usefulness, speed, accuracy, overall quality, perceived effort, boredom.

Source: Matamala, Romero-Fresco & Daniluk (forthcoming)

# SR in VO: results

- Speed: manual, respeaking, ASR revision.
- Accuracy: manual, ASR revision, respeaking.
- Respeaking allowed the highest number of participants to finish.
- Respeaking obtains better scores in perceived effort and boredom.
- Manual transcript, better scores in accuracy and overall quality.
- Respeaking: impressed with many possibilities (combination of techniques?), more job satisfaction but need for specific training and further research.

# Machine translation in voice-over

- 12 AVT MA students: human translation vs post-editing of a wildlife documentary excerpt (En>Es)
- Temporal, technical, cognitive effort (translation vs post-editing) using InputLog
- Quality assessment by 6 experts, dubbing studio, and 56 end users

Source: Ortiz-Boix & Matamala (forthcoming)

# MT in VO

- Results obtained for: both excerpts, excerpt 1, excerpt 2
- Temporal effort: post-editing faster although results are only statistically significant in excerpt 1
- Technical and cognitive effort: post-editing requires less effort, although differences are only statistically significant globally and in excerpt 1
- Translation quality is slightly higher in all conditions but results are not statistically significant and there are differences according to excerpts/groups.

# Speech synthesis in voice-over

- On-going



# Speech recognition in audio description

- Soundtrack extraction, speech activity detection, speaker diarization, and speech-to-text transcription
- Film “Closer” (English, Catalan)
- DER and WER measures: low performance because of training conditions of the system and employed AD materials.

Source: Delgado, Matamala & Serrano (forthcoming)

# Machine translation in audio description

- Pre-test to select the MT engine (Fernández-Torné & Matamala 2014): Google Translate.
- Main experiment: 12 AVT MA students: human creation/translation/post-editing (En>Cat)
- Temporal, technical and cognitive effort (InputLog)
- Perceived effort (questionnaires)

# MT in AD: results

- Temporal effort: no statistical differences.
- Technical effort: AD creation and AD translation, more keyboard action than post-editing. AD translation and post-editing, greater number of mouse scrolls than AD creation.
- Cognitive effort (pause-to-word ratio): statistically higher in AD creation.

# Speech synthesis in audio description

- Pre-test to select the “best” human/natural male/female voices
- Main experiment: 67 blind and visually-impaired volunteers
- Questionnaire inspired by ITU (1994)

Source: Fernández-Torné & Matamala (2015)

# SS in AD: results

- Natural voices have higher values than artificial voices
- No statistical differences between male/female natural voices
- Overall impression/acceptance: mean higher than 3.2 (on a 5-point scale)
- 94% participants: TTS is an alternative acceptable solution although not the preferred one

# Conclusions

- Small-scale exploratory project: more research needed with wider samples and more excerpts
- Need for customization of technologies
- Not only time (productivity) but also opinions of professionals and end users

# Speech and translation technologies for voice-over and audio description: final results of the ALST project

**Anna Matamala (Universitat Autònoma de Barcelona)**

**[anna.matamala@uab.cat](mailto:anna.matamala@uab.cat)**

**6th International Conference Media for All  
Western Sydney University, 17-18/09/2015.**

**Funding: FFI-2012-31024, 2014SGR0027**

