

Languages & The Media 2018 - 12th International Conference
on Language Transfer in Audiovisual Media

More than words Voice in audio description

Anna Matamala, María Machuca, Antonio Ríos
anna.matamala@uab.cat

TransMedia Catalonia Research Group

Departament de Traducció, d'Interpretació i d'Estudis de l'Àsia Oriental
Universitat Autònoma de Barcelona



Universitat Autònoma
de Barcelona

TRANS
MEDIA
CATALONIA

Rationale

AD research: focus on content selection and AD writing.

Voice has been researched in relation to technology (TTS AD), but what about prosodic features?

Aims

To revise recommendations related to voice in different guidelines.

To analyse the prosodic features of a corpus of AD.

To identify user preferences in terms of voicing.

The literature says...

Fryer (2016: 88): describers have been traditionally “encouraged to use a particular neutral way of speaking” and a “neutral delivery has come to be recognized as the norm”. However, specific features of each production should be considered.

Snyder (2014): vocal skills, one of the four fundamental elements of AD. “We make meaning with our voices”, “voicer’s delivery should be consonant with the nature of the material being described”

Research shows...

Cabeza-Cáceres (2013):

- Spanish and German: “uniform”
- British AD: “adapted”
- USA AD: “emphatic”.

None of these prosodic styles seem to impact on comprehension, and user preferences do not always coincide: same percentage like and reject uniform and emphatic intonations (ibid: 331).

Guidelines say...

Spanish guidelines: clear voice and voicing must be neutral with appropriate intonation, rhythm and vocalisation.

ISO: good native skills and ability to articulate, the voice should complement the content being described, and acknowledges different styles can be used.

ADLAB: voice qualities match the film's genre and style. "Little or no research on which voices fit which film genre best", but generally based on contrast with the film voices.

Guidelines say...

Ireland: neutral voice, emotion can be added at certain points.

German: neutral voicing, but change in humorous productions?

Guidelines say...

French: voice adapted to the emotion of the scene but shall keep certain neutrality.

ITC: clear, pleasant and expressive voice, unobtrusive and neutral, not lifeless or monotonous.

Poland: not distract users, relation to nature of the programme, neutral and discreet, not monotonous.
etc.

Summary

Voice is a key element in AD.

“Neutral”, “adapted”... what does it mean?

Our research

1. Prosodic analysis of a corpus of AD in Spanish.
2. Analyse user perception

The corpus

VIW corpus

<http://pagines.uab.cat/viw/>

10 professional Spanish ADs based on the same input:
compare the result in terms of prosodic features?

<http://pagines.uab.cat/viw/content/spanish-audio-descriptions>

The Project

The Film: WHW

The Corpus

Researchers

Contribute

Research

Contact

The Corpus

The VIW corpus is made up of audio descriptions in different languages of the same short film, and it incorporate a **search engine** (forthcoming). It currently includes 10 professional audio descriptions in English, 10 professional audio descriptions in Spanish and 10 professional audio descriptions in Catalan. All 30 audio descriptions were commissioned to professionals, who were remunerated. It also includes 10 audio descriptions in Spanish and 7 audio descriptions in Catalan made by students (text only, not recorded audio/video files). We hope to expand it in the near future with more audio descriptions. You are invited to read the section "Contribute".

Audio and video files with audio description by professional describers

 English  Catalan  Spanish

Text files of audio descriptions by students

 Catalan  Spanish

ACCESS TO THE CORPUS SEARCH ENGINE

All materials are also available at UAB's open access repository: <https://ddd.uab.cat/record/147267>.

SPANISH AUDIO DESCRIPTIONS



English audio descriptions



Catalan audio descriptions



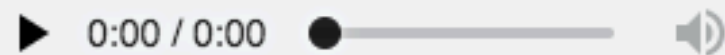
Spanish audio descriptions

All materials are also available at UAB open access repository: <https://ddd.uab.cat/record/147267>.

Aptent:



Aristia:



CEIAF:



List of available corpora

You can also use the [BROWSE facility](#) to look for a specific file.

- **WHW-EN-Pr:** The WHW-EN-Pr corpus contains professional audio descriptions in English for the short movie "What Happens While". [access](#)
- **WHW-ES-Pr:** The WHW-ES-Pr corpus contains professional audio descriptions in Spanish for the short movie "What Happens While". [access](#)
- **WHW-CA-Pr:** The WHW-CA-Pr corpus contains professional audio descriptions in Catalan for the short movie "What Happens While". [access](#)
- **WHW-ES-St:** The WHW-ES-St corpus contains audio descriptions in Spanish for the short movie "What Happens While" made by students. [access](#)
- **WHW-CA-St:** The WHW-CA-St corpus contains audio descriptions in Catalan for the short movie "What Happens While" made by students. [access](#)

<http://transmediacatalonia.uab.cat/web/>

The methodology

Audio files of all 10 AD, segmented into intonation groups.

Analysed with PRAAT.

Duration, Pitch, and Amplitude, which are indicators of speech rate, voice quality, volume and loudness of the sound.

Service providers: Aptent, Aristia, CEIAF, Edsol, Ericsson, Kaleidoscope, Navarra de Cine, SDI, Soni2, Trágora.

Results

50% female, 50% male.

Prosodic values available (please email us).

Examples to be shown on the presentation.

Results

- Aristia: lowest pitch, high volume, low speech rate.
- Aptent: highest pitch and volume, speech rate slightly above average.
- Kaleidoscope: low pitch, low volume and fast speech rate.
- Edsol: low pitch, average volume, fast speech rate.
- Ericsson: high pitch, low volume, lowest speech rate.
- Sonidos: highest pitch, highest speech rate, fastest speech rate.

User perception: the test

60 participants (29 sighted, 31 with sight loss).

Online lab test: not in context (future research).

The most different voices were chosen: 3 male, 3 female.

Part 1: Participants listened to short excerpts with 18 pairs of voices and selected the most pleasant voice.

Part 2: neutrality. Not discussed here.

Results

FEM: Sp_1 (Aptent), Sp_2 (Aristia), Sp_3 (Kaleidoscope).

Sp2, preferred. Sp_3, less pleasant (low pitch and volume and high speech rate).

Sp1 and Sp2: no statistical differences. Sp1 preferred by sighted users and Sp2 by users with sight loss.

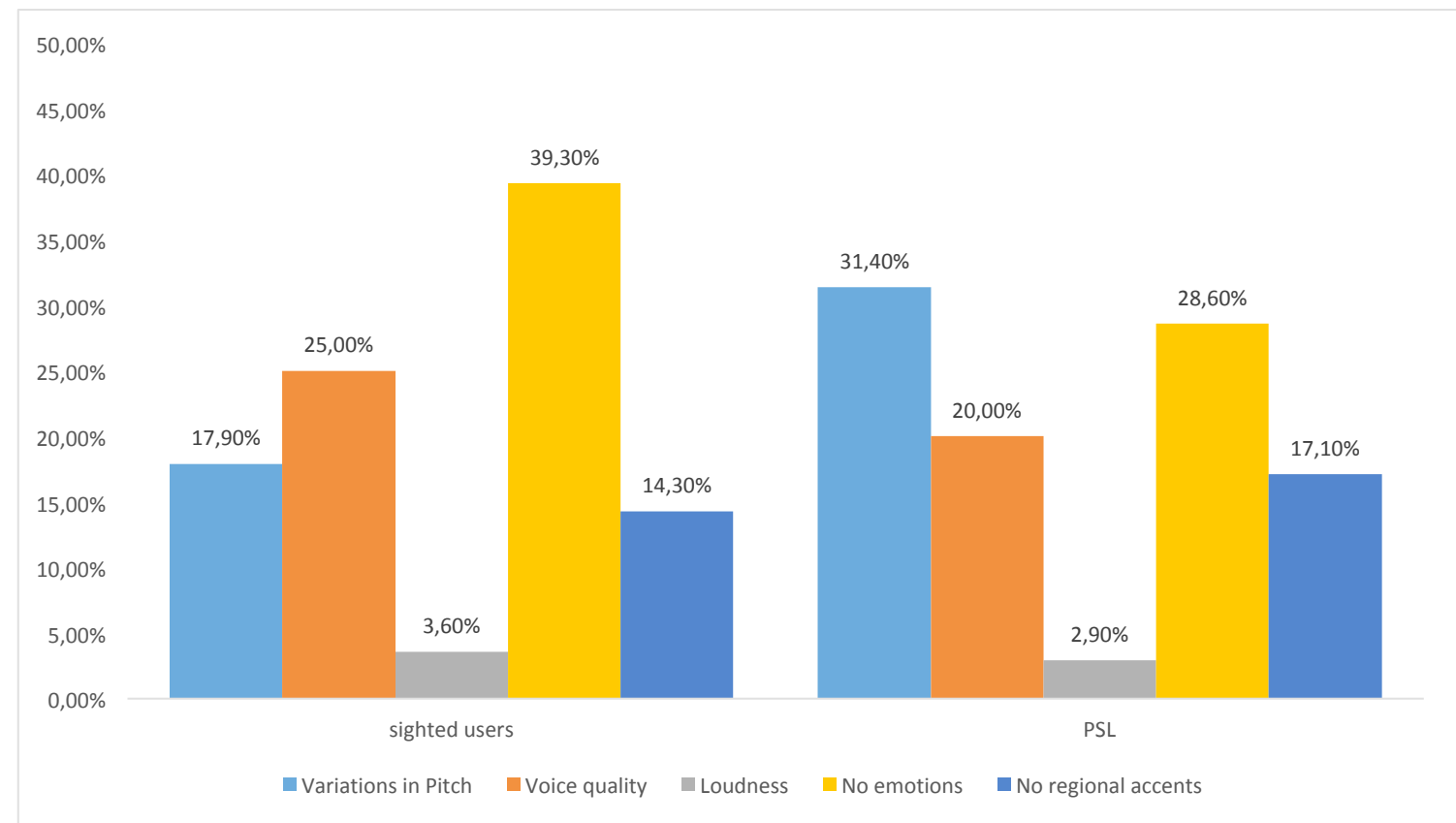
Results

MALE: Sp_1 (Edsol), Sp_2 (Ericsson), Sp_3 (Sonidos).

No differences sighted versus persons with sight loss.
More pleasant: Speaker 2.

Next steps

Neutrality and user perception: ongoing analysis.



Next steps

Neutrality and user perception.

Differences among languages?

Advance in prosodic analysis, beyond impressionistic descriptions.

Disclaimer

TransMedia Catalonia is a research group funded by Secretaria d'Universitats i Recerca del Departament d'Empresa i Coneixement de la Generalitat de Catalunya, under the SGR funding scheme (ref. code 2017SGR113).

The project **NEA** has received funding from Ministerio de Economía y Competitividad, grant agreement No FFI2015-64038-P, MINECO/FEDER, UE.



Unión Europea

Fondo Europeo
de Desarrollo Regional
"Una manera de hacer Europa"

Disclaimer

Texts, marks, logos, names, graphics, images, photographs, illustrations, artwork, audio clips, video clips, and software copyrighted by their respective owners are used on these slides for personal, educational and non-commercial purposes only. Use of any copyrighted material is not authorized without the written consent of the copyright holder. Every effort has been made to respect the copyrights of other parties. If you believe that your copyright has been misused, please direct your correspondence to anna.matamala@uab.cat stating your position and we shall endeavour to correct any misuse as early as possible.

This document and its contents reflect the views only of the authors. TransMedia Catalonia and the funding organisms referred to in this presentation cannot be held responsible for any use which may be made of the information contained therein.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.

Languages & The Media 2018 - 12th International Conference
on Language Transfer in Audiovisual Media

More than words Voice in audio description

Anna Matamala, María Machuca, Antonio Ríos
anna.matamala@uab.cat

TransMedia Catalonia Research Group

Departament de Traducció, d'Interpretació i d'Estudis de l'Àsia Oriental
Universitat Autònoma de Barcelona



Universitat Autònoma
de Barcelona

TRANS
MEDIA
CATALONIA