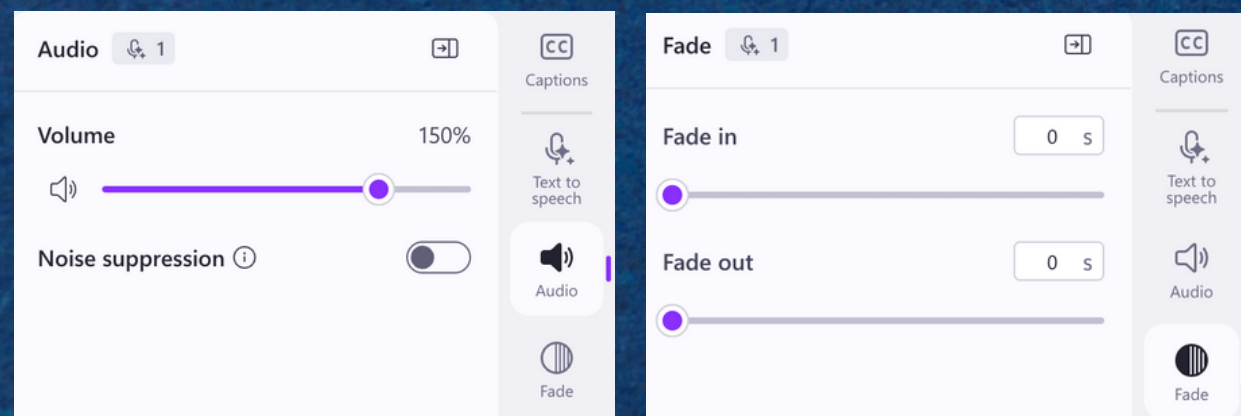# Testing Clipchamp as an Automated Text-to-Speech Recognition Software for Audio Description Research

José Luis Castillo-Flores
luis.castillo@ulpgc.es
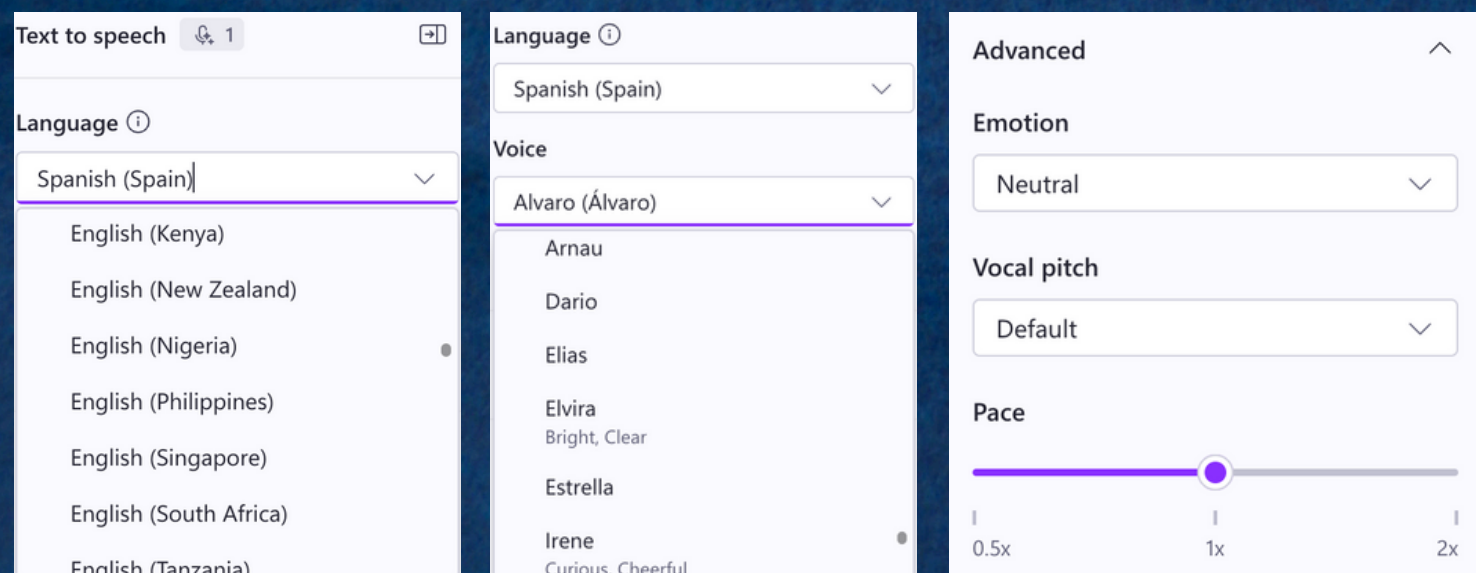ARSAD | March 20, 2025

Sound editing options

- Volume adjustment (original track vs. AD track)
- Background noise suppression
- Sound fading
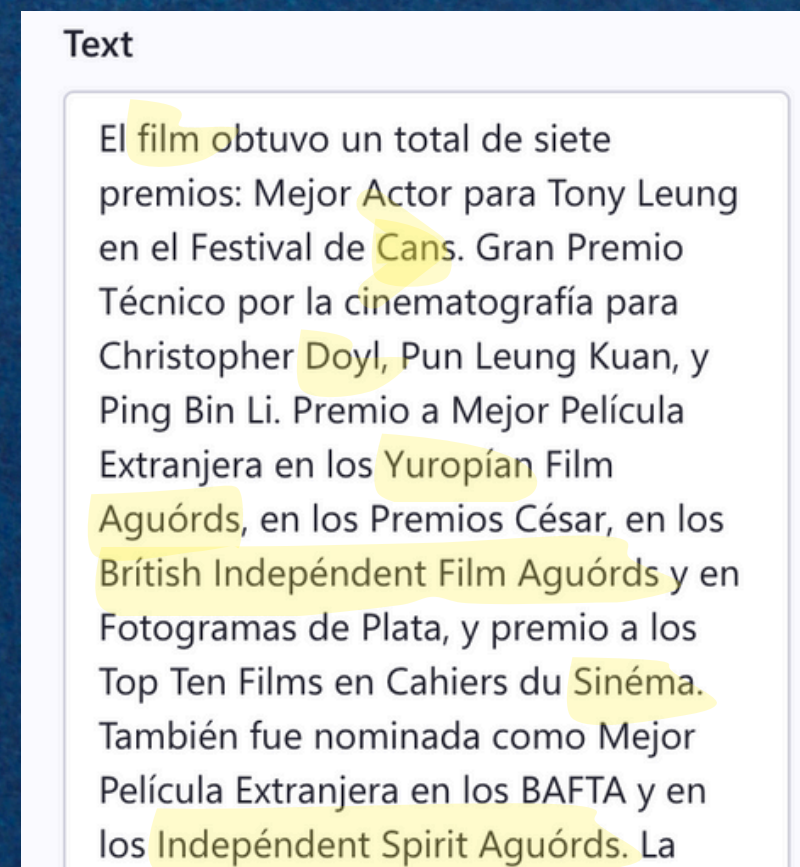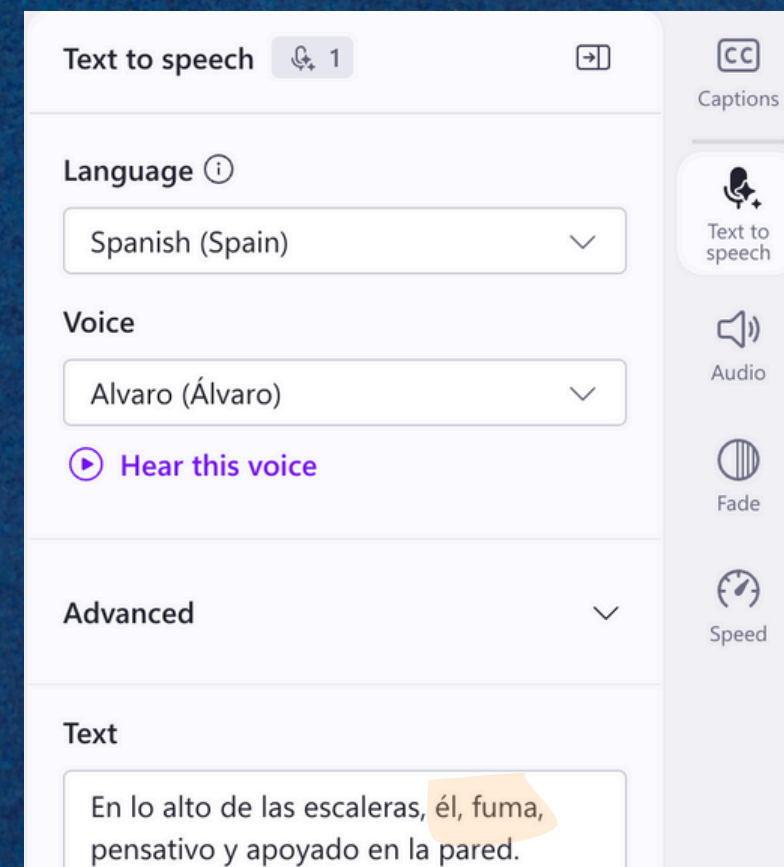- Text-to-speech recognition feature

Strengths

- Multiple languages and accents
- Vast array of voices
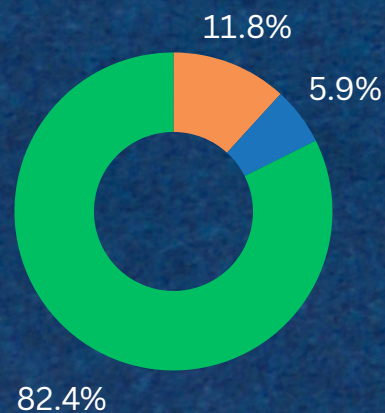- Type of emotion and of vocal pitch
- Vocal pace adjustment

Weaknesses

- Intonation flops for specific chains of words or long segments
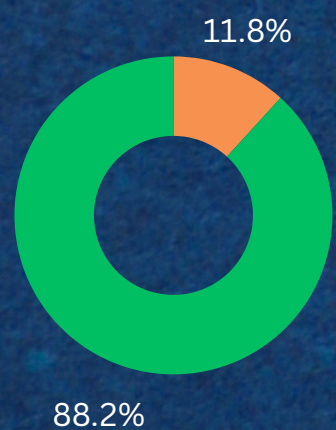- Cannot recognize multiple languages in a text

Experiment

- 8 clips ~ 45 minutes
- 17 BPS participants
- No previous information on how AD has been voiced
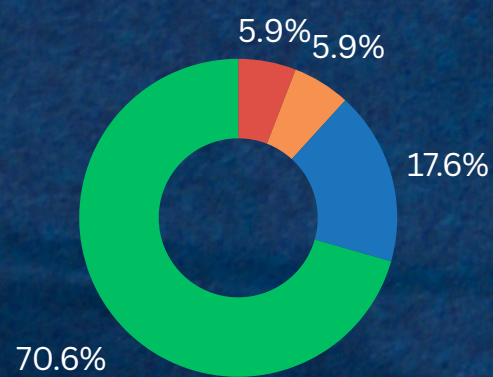- Likert scale questionnaire
- Group discussion

Q1: The quality of the voice was appropriate

In this section, 2 subjects **slightly disagreed** with the statement (11.8%). In contrast, 1 **strongly agreed** (5.9%), and 14 **completely agreed** (82.3%). In total, 15 participants believed that the quality of the voice was appropriate (88.2%).
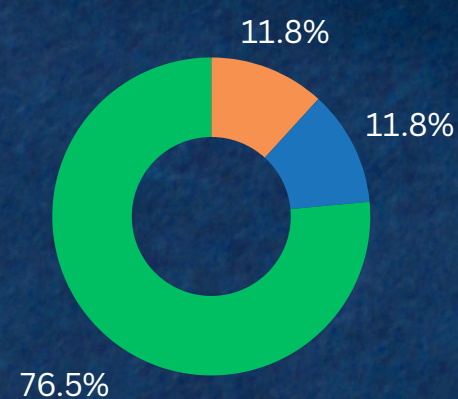
11.8%
5.9%
82.4%

Q2: The intonation was appropriate

This time, 2 subjects **slightly disagreed** with the previous statement (11.8%), while the other 15 participants **completely agreed** (88.2%).
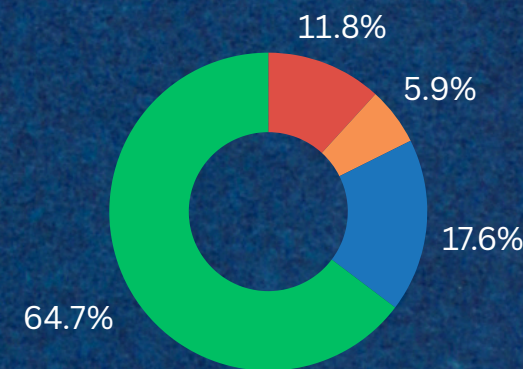
11.8%
88.2%

Q3: Speech rate was appropriate

Here, 1 subject **completely disagreed** (5.9%), and 1 **slightly disagreed** (5.9%). Nonetheless, 3 **strongly agreed** (17.6%) and 12 **completely agreed** (70.6%). In general, 15 participants found the speech rate appropriate (88.2%).
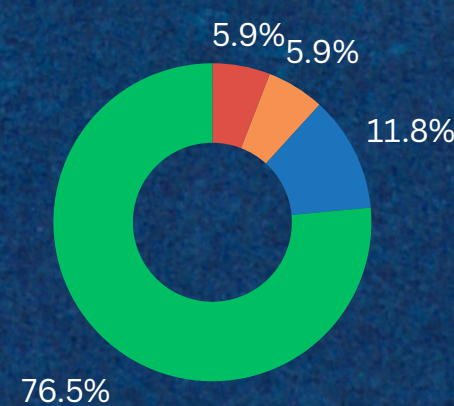
5.9% 5.9%
17.6%
70.6%

Q4: Pauses between sentences were appropriate

On this occasion, 2 subjects **slightly disagreed** (11.8%). Nevertheless, 2 **strongly agreed** (11.8%) and 13 **completely agreed** (76.5%). Altogether, 15 subjects expressed that the pauses between sentences were appropriate (88.3%).
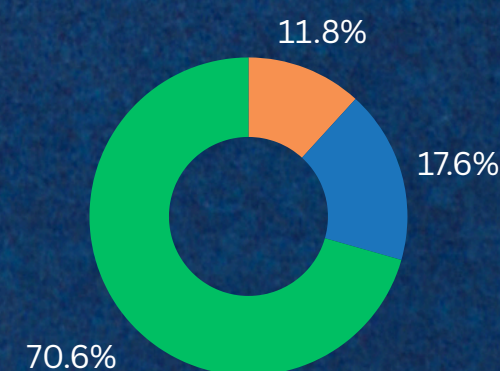
11.8%
11.8%
76.5%

Q5: The voice was natural

On the one hand, 2 subjects **completely disagreed** (11.8%) and 1 **slightly disagreed** (5.9%). On the other, 3 subjects **strongly agreed** (17.6%) and 11 **completely agreed** (64.7%). Overall, 14 people found the voice natural (82.3%).
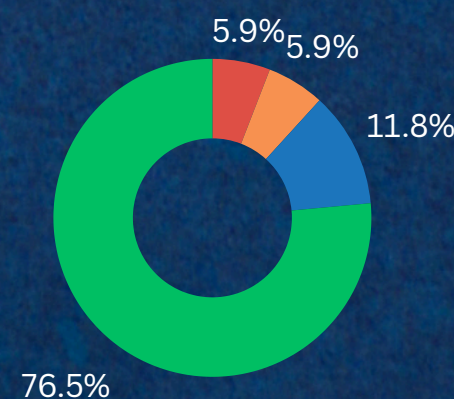
11.8%
5.9%
17.6%
64.7%

Q6: The voice was pleasant

In this case, 1 subject **completely disagreed** with the statement (5.9%), and 1 **slightly disagreed** (5.9%). However, 2 subjects **strongly agreed** (11.8%), and 13 **completely agreed** (76.5%). On the whole, 15 people believed the voice was pleasant (88.3%).

5.9% 5.9%
11.8%
76.5%

Q7: My level of satisfaction with the voice is adequate

In this question 2 subjects **slightly disagreed** (11.8%), 3 **strongly agreed** (17.6%), and 12 **completely agreed** (70.6%). In sum, 15 participants were satisfied with the voice (88.2%).

11.8%
17.6%
70.6%

Q8: The voice is suitable for future audio descriptions

In this last section, 1 subject **completely disagreed** with the statement (5.9%), and another **slightly disagreed** (5.9%). Contrarily, 2 subjects **strongly agreed** (11.8%), and 13 **completely agreed** (76.5%). All in all, 15 participants believed that the voice is suitable for future audio descriptions (88.3%).

5.9% 5.9%
11.8%
76.5%

Thank you!

José Luis Castillo-Flores
luis.castillo@ulpgc.es
ARSAD | March 20, 2025