ALEXANDER KURCH, CHRISTIAN DAVID

TEXT-TO-SPEECH AUDIO DESCRIPTION:
POTENTIALS OF WEB-BASED APPLICATIONS
TOWARDS MORE ACCESSIBILITY

Vitae

Alexander Kurch

- 5 years of speech-to-text-interpreting for the hard of-hearing
- focus: audiovisual translation, media accessibility, text technologies

Christian David

- 20 years of software development (web, apps, exhibits)
- co-founder of VIDEO TO VOICE, founded in 2016
- focus: developing web-based tools to produce and deliver accessible content

Overview

- **text-to-speech audio description (TTS AD)**

- **research overview**

- **production workflow**

- **ways of delivery and the web**

Text-to-Speech (TTS)

What is text-to-speech synthesis?

- written text input automatically transformed into spoken language

- with an artificial voice via computer software

- modern TTS:

- "produce an effect far more natural than a few years ago." (Szwarkowska 2011)

Text-to-Speech Audio Description (TTS AD)

Why using TTS for audio description?

- people with visual impairments use speech synthesis for internet access and assistive computer technology at home and at work (Möbius/Haiber 2010)

- assumption: acquaintance and familiarity with TTS

Text-to-Speech Audio Description (TTS AD)

Why using TTS for audio description?

- AD script can be read out by TTS systems

- "text-to-speech audio description (TTS AD) is proposed here in order

- to increase the AD output and to make AD more available"(Szarkowska 2011)

- Saving costs for studio recording with voice artists

- Saving costs for mastering AD parts and film material

Text-to-Speech Audio Description (TTS AD)

Why using TTS for audio description?

- modern TTS engines use deep learning strategies to optimize „themselves"

- global players like Amazon, Google or Microsoft provide access to their TTS engines as part of their cloud architecture → affordable for everyone

Research

Studies in Poland

2011: complete feature film (n = 24) (Szarkowska 2011)

- 58 % for TTS AD as a permanent solution

2012 complete foreign film (n = 20) (Szarkowska, Jankowska 2012)

- 70 % for TTS AD as a permanent solution

Research

Study in Spain (Fernández-Torné/Matamala 2015: 63)

- material: 3 minute clips (feature film), n = 67
- 94 % as an alternative solution

Study in Germany (Kurch 2015/2016)

- material: 5 minute clip (documentary), n = 18
- 72 % as alternative solution

Text-to-Speech Audio Description (TTS AD)

Desiderata for production workflow:

- one application: merging video player, TTS software and authoring functions

- timecoding of AD slots and automatic calculation of captions per frame

- simultaneous preview: video material and TTS AD audio track

- delivery: easy file export from producer > playback on end-user devices

Frazier: web based AD authoring

- web-based video player

- create audio descriptions online

- live preview

- generate TTS audio track

- export for 3rd party tools (.WAV, .VTT)

ablePlayer: accessible media player

- https://ableplayer.github.io/ableplayer/

- fully accessible web based media player

- AD playback using screen reader and exported .VTT file

flexMOTE: play AD on smartphone

- web-based video player on demo website

- web-based audio player on smartphone

- connect through QR code

- automatically keeps AD in sync w/ video player (play, pause, seek)

Conclusion

- the web is „everywhere", everything is connected

- with web technology we reach a wide range of devices (pc, laptop, smartphone, tablets) with standardized accessibility features

- new technologies like text-to-speech & speech recognition allow integrated and highly automated workflows to produce AD

- QR code, NFC, beacons & audio fingerprinting can help to deliver audio description to „any" device.

Bibliography

Fernández i Torné, Anna; Matamala, Anna (2015): Text-to-speech vs. human voiced audio descriptions: a reception study in films dubbed into Catalan. In: *JoSTrans - Journal of Specialised Translation* (24), 61-88. www.jostrans.org/issue24/art_fernandez.pdf [30.08.2017]

Kurch, Alexander (2016): Audiodeskription mit Sprachsynthese - Eine Perzeptions- und Rezeptionsstudie am Beispiel des Dokumentarfilms „buy buy st. pauli". University of Hildesheim : unpublished MA thesis.

Möbius, Bernd; Haiber, Udo (2010): Verarbeitung gesprochener Sprache". In: Carstensen, Kai-Uwe / Ebert, Christian / Ebert, Cornelia / Jekat, Susanne J. / Klabunde, Ralf / Langer, Hagen (Hrsg.) (2010): *Computerlinguistik und Sprachtechnologie - Eine Einführung*. Heidelberg: Spektrum, Akad. Verl, 214-235

Bibliography

Szarkowska Agnieszka (2011) "Text-to-speech audio description. Towards a wider availabilty of AD. In: *Journal of Specialised Translation* in January 2011. http://www.jostrans.org/issue15/art_szarkowska.php [30.08.2017]

Szarkowska, Agnieska / Anna Jankowska (2012). "Text-to-speech audio description of voice-over films. A case study of audio described *Volver* in Polish." Elisa Perego (Hrsg.) (2012). *Emerging topics in translation: Audio description*. Trieste, Italy: EUT Edizioni Università di Trieste, 81-98.

Thank you for your attention

c.david@videotovoice.com

alexander.kurch@web.de