Software-based Audio Description with Frazier

Marina Gerster

IUED Institut für Übersetzen und Dolmetschen
ZHAW, Winterthur
`marina.gerster@zhaw.ch`

Abstract

This article [25] gives insight into how audio described films with synthetic audio description voices are received by blind and visually impaired people. A survey was conducted with the target audience to evaluate the quality and identify the future prospects of synthetic audio description voices. Results show that synthetic voices offer the opportunity to make audio description more widely available.

1   Introduction

Against the background of the UN Convention on the Rights of Persons with Disabilities (UN 2006), and within the paradigm of social inclusion and participation, audio description (AD) enables blind and visually impaired people to experience a variety of visual media (Benecke 2014).

The conventional method for the production of audio descriptions – i.e. in a team of three people and without any supporting software, as employed by many television companies – is extremely costly and time-consuming (Jekat/Oláh 2016). This hinders the audio description of large volumes of (internet) videos.

Current technology allows to produce audio description by means of synthetic voices. *Frazier* is a web-based software application with integrated text-to-speech technology.

The main goal of this study is to develop the unique selling proposition of *Frazier*'s beta-version. Furthermore, it is investigated whether blind and visually impaired people consider synthetic AD voices as an acceptable alternative to conventional human AD voices.

In particular, this work aims to answer the following questions:

-        What is the target audience's assessment of the quality of synthetic AD voices?

-        Can a preference be identified for either the synthetic or human AD voice?

-        Does the knowledge transfer in an audio described film with synthetic AD voices take place in the same way as in an audio described film with human AD voices?

-        What are the future prospects of audio described films with synthetic AD voices?

2   Frazier

Frazier was developed by the Berlin start-up VIDEO TO VOICE GmbH. *Frazier* is a web-based software application for AD production which employs synthetic voices for the immediate implementation of the descriptions. Its beta-version is not yet available to the public; I was granted permission to test it for research purposes.

VIDEO TO VOICE claims that *Frazier* allows a simple, fast and efficient AD production (VIDEO TO VOICE o. J.). The user-friendliness of the software was confirmed in this study. The production process was intuitive and simple to implement.

3   Method

In order to address the above-mentioned research questions, the synthetic AD of the short film *Wie immer* (Sethna 2010) was created with *Frazier*. This allowed to make a comparison between the synthetic and human AD of the same film. In addition, a survey was conducted with the target group (blind and visually impaired people) and a comparison group (sighted people).

Half of the respondents were first shown the audio described film with synthetic AD voices and, subsequently, the audio described film with human AD voices. The other half was presented with the two AD versions in reverse order. After each listening session, the groups were asked to answer questions regarding the quality of AD voices, the film content, and to make a comparison between the two AD versions.

4 Results and discussion

4.1 Quality

The majority of blind and visually impaired people was pleasantly surprised by the good quality of the synthetic AD voices and by the considerable progress that this technology has made in the recent past in terms of sound, intonation and prosody, emphasis on individual words and fluency of speech. The target group predominantly rated the synthetic AD voices as positive.

4.2 Preference

The target group tended to prefer the human AD voice of the audio described film Wie immer (Sethna 2010). Especially in films where the listening experience is of primary importance, the target group prefers a human AD voice. However, synthetic AD voices will certainly go through further technical development and, according to the target group, will also be conceivable in entertainment films in the future.

4.3 Knowledge transfer

The results of the content questions showed that both the synthetic and human AD voices performed equally in terms of knowledge transfer, i.e. they had a comparable impact on the respondents' comprehension of the film.

4.4 Future potential

This study suggests that synthetic AD can certainly be an alternative solution to human AD, as well as an opportunity to make AD more widely available while, at the same time, saving on production costs. The range of audio described videos can thus be easily extended to include, in particular, Internet or YouTube videos. In addition, the use of synthetic AD voices is suitable in educational contexts, where synthetic AD allows lecturers and/or students to produce their own barrier-free films with a minimum of time and expense.

Overall, synthetic AD represents a significant contribution towards the accessibility of teaching materials for blind and visually impaired people.

5 Conclusion

This study indicates that the human voice will not be replaced completely by synthetic AD voice in the near future. A human voice is preferred by the target group especially for films where the listening experience is the first priority. The use of synthetic AD voices seems particularly suitable for educational films or YouTube videos. However, this technology will certainly advance over the coming years and it is conceivable that it will also be used in feature films in the future.

References

Benecke, Bernd. 2014. *Audiodeskription als partielle Translation: Modell und Methode.* Berlin: Lit.

Jekat, Susanne, and Annegret Oláh. 2016. Theorie und Methode der Audiodeskription: ein Pilotprojekt. In N. Mälzer (Hrsg), *Barrierefreie Kommunikation – Perspektiven aus Theorie und Praxis*, Berlin: Frank & Timme 69-94.

Sethna, Zubin. 2010. *Wie immer*. Germany. Short film, 06:38. Screenplay by Dorothea Nölle. https://vimeo.com/72959210 [zit. 12.07.2018].

UN. 2006. *United Nations Convention on the Rights of Persons with Disabilities*. United Nations, http://www.un.org/disabilities/documents/convention/convention_ac-cessible_pdf.pdf [zit. 09.07.2018].

VIDEO TO VOICE. o. J. *Frazier. Audiodeskription mit Text-to-Speech-Technologie* (TTS-AD). https://www.videotovoice.com/startseite-alt/ [zit. 09.07.2018].