JoSTrans
The Journal of Specialised Translation

www.jostrans.org · ISSN: 1740-367X

Szarkowska, A. (2011). Text-to-speech audio description. Towards wider availability of AD. *The Journal of Specialised Translation, 15*, 142-162. https://doi.org/10.26034/cm.jostrans.2011.509

Text-to-speech audio description: towards wider availability of AD

Agnieszka Szarkowska, University of Warsaw

ABSTRACT

This paper addresses the feasibility of text-to-speech audio description (TTS AD)[1] and its reception among visually impaired people. First, a new method of producing AD to be read by speech synthesis software is proposed. The method is then put to the test by examining a feature film *The Day of the Wacko*, screened with TTS AD. Finally, the results of the reception study are presented and discussed, followed by recommendations and suggestions for further research. The results of the TTS AD study – the first of its kind – demonstrate an untapped potential of the method. The majority of the blind and partially sighted respondents participating in the screening and survey declared they accepted TTS AD as an interim solution, while many others were also in favour of TTS AD becoming a permanent option.

KEYWORDS

Audio description, text-to-speech audio description, speech synthesis, audiovisual translation, accessibility, blind, visually impaired

Introduction

In recent years we have witnessed a proliferation of efforts aimed at making audiovisual programmes accessible to people with visual impairments through audio description (AD) (see Orero 2007). This flurry of activity could be observed in many areas. First, there are grassroots initiatives of the visually impaired themselves (Greening and Rolph 2007, Ciborowski 2008). Second, broadcasters and professionals have also taken action (Benecke 2004, Benecke 2007). More and more legal regulations in relation to media accessibility are being passed now, e.g. the Audiovisual Media Directive 2007/65/EC which requires implementation in the various EU Member States. Higher education institutions have also been involved in AD development—both in terms of academic research (Schmeidler and Kirchner 2001, Braun 2007, Vercauteren 2007, Salway 2007, Chmiel and Mazur 2008) and audio describer training (Remael and Vercauteren 2007, Snyder 2008). Not only are audio described films made available in cinemas, on television, DVD/Blu-ray and over the Internet, but projects aimed at introducing audio description to theatres, opera (Matamala and Orero 2007), museums and art galleries have proliferated exponentially in the last two decades.

What is audio description?

From the viewpoint of disability studies, audio description is "an enabling service" (Holland 2009: 171) which promotes inclusion by making it possible

for blind and partially sighted viewers to gain access to film, theatre and opera. In Europe, AD has been readily embraced by translation studies researchers under the umbrella of *accessibility*, understood as "making an audiovisual programme available to people that otherwise could not have access to it" (Díaz Cintas 2005: 5).

Just as is the case with the term *translation*, *audio description*[2] can refer to both product and process. The former stands for a special type of narration directed at spectators who are blind or visually impaired. This narration is inserted within existing pauses in dialogue. AD is an audio account of the visual and aural content important to understand audiovisual material; AD can include information in relation to actions, scene changes, on-screen text, descriptions of characters, their movements and body language, explanation of sound effects, etc. (see Ofcom 2000, Vercauteren 2007). Audio description as a process concerns the production of this type of narration.

A lengthy preparation process and high production costs are among the greatest obstacles to the wider availability of audio description. In Ofcom's (2000: 12) estimates, "a two-hour film may take up to sixty hours to prepare [and] on average it takes one describer a working week to produce between one and a half and two hours of described programming." The standard AD production process consists of the following stages: selecting the material to be audio described, viewing the programme, preparing and reviewing the script, recording the description and mixing it with the main programme audio in the case of pre-recorded AD or rehearsing the script with the video in the case of AD delivered live. The reviewing stage can include consultations with a visually impaired person in order to make sure the content of the script is clear and understandable. The voicing of AD can be done by audio describers themselves or be commissioned to actors or voice talents. Professionally prepared, audio description is usually the sum of work carried out by a team of individuals. In view of the above, audio description is often considered a "cost" service "which broadcasters must provide in order to comply with governmental and broadcast mandates" (Udo and Fels 2009). Added to this is the fact that AD is sometimes perceived as "catering to the needs of very small and specific population" (*ibid*.). All this significantly hampers the marketability and wider availability of audio description.

Even though AD is a relatively new field, certain standards have already been established. Among the most frequently quoted are the three golden rules listed in the *ITC Guidance on Standards for Audio Description* (2000: 9): "describe what is there, do not give a personal version of what is there and never talk over dialogue or commentary." Snyder (2008: 195) proposes an acronym WYSIWYS, standing for 'What You See Is What You Say,' to emphasize that describers are supposed to "objectively recount visual aspects

of an [...] audiovisual programme." Essentially, most audio description standards currently in use stress neutrality, objectivity and impartiality.

The practice of conventional audio description has recently come to be questioned by a number of experimental studies. For instance, Fels *et al*. (2005) have suggested an alternative approach to AD, marked by the integration of AD in the film production process and a departure from the omnipresent third-person narrator providing an "objective," factual and interpretation-free description to a subjective first-person-AD-story-teller. Another interesting study was that of Udo and Fels (2009b), in which they investigated the reception of the audio description of a live theatre performance of Shakespeare's *Hamlet*, where the AD was "written in iambic pentameter and delivered from Horatio's point of view." The same Canadian researchers carried out an analysis of the live AD of a fashion show (2009a). All of the above clearly demonstrates further scope for experimentation in the area of audio description.

In view of the above, it can be seen that—despite the growing enthusiasm of AD researchers, broadcasters and the visually impaired community—audio description is still a product/service whose availability leaves a lot to be desired, mainly owing to the complex and costly production process and restricted distribution. The number of audiovisual products available on DVDs/Blu-ray is still insufficient, while screenings with live AD are few and far between and reach only a handful of spectators.

Text-to-speech audio description

With these shortcomings of conventional audio description in mind, text-to-speech audio description (TTS AD) is proposed here in order to increase the AD output and to make AD more available. The idea behind it is that instead of recording a human voice reading out the AD script, TTS AD can be read by speech synthesis software. Modern text-to-speech applications convert text input into a speech waveform with the use of special algorithms (Cryer and Home 2008: 5), producing an effect far more natural than speech synthesisers did a few years ago.

Figure 1 below shows how TTS AD works. First, the AD script is created and then inserted into a film between dialogues as "AD subtitles," using subtitling software. What this means in practice is that the AD script is cut into chunks, each of them having allocated time-codes, i.e. the in and out times indicating when the description should begin and end. Paired with voice recognition capability offered by professional commercially available subtitling software, TTS AD allows for extreme precision in inserting the chunks within pauses in dialogue. The resulting text file with synchronised time-codes is then read by

speech synthesis software while the audiovisual material is simultaneously played by a film player. After the preparation of "AD subtitles" comes the reviewing stage. Ideally, the script should be developed in consultation with a visually impaired person.

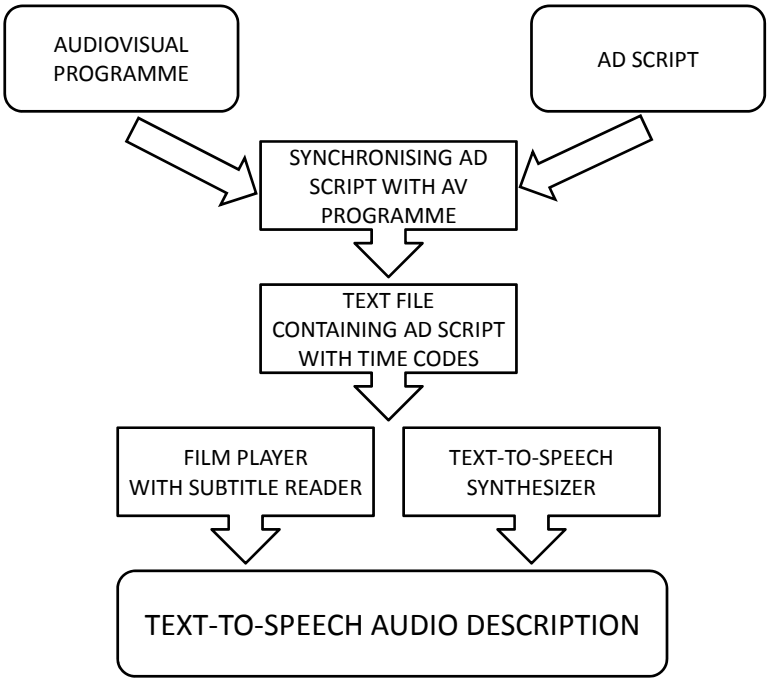

Fig. 1 Text-to-speech audio description preparation process

In order to watch the film with text-to-speech audio description, a film player with a subtitle reader is required. For the experiment described herein (see below), I used the freeware programme BestPlayer (version 2.106) combined with a female radio quality Polish synthetic voice (at a sampling rate of 22 kHz) named Ewa, from the text-to-speech application Ivona Reader (manufactured by Ivo Software). The player allows the viewer to set the reading speed and volume parameters.

Text-to-speech audio description has several advantages. From the perspective of the audio description provider, TTS AD offers unequalled cost-effectiveness in terms of AD production in comparison with conventional methods of producing audio description. TTS AD does not require the recording of the AD script (for pre-recorded AD), nor does it incur any human labour costs for the reading out of the AD script (for live AD). Furthermore, in contrast to audio describers involved in the production of conventional AD, who need to be able to develop "the vocal instrument through work with speech and oral interpretation fundamentals" (Snyder 2008: 196), audio describers for TTS AD

do not need to have any particular vocal skills.

From the end user's point of view, TTS AD also helps many blind and partially sighted people save money, as they already have speech synthesis software at home or at work and are accustomed to using it in their daily lives. Thanks to the high quality of speech synthesis software available now in many languages, watching a film with synthetic AD can be an enjoyable and entertaining experience.[3] For those watching the audiovisual programme on the net,[4] another advantage is that the solution does not require access to a high-speed Internet connection because the viewer is simply offered a text file with the AD script (in .txt or .sub format) to be read out by a text-to-speech programme. The solution seems particularly attractive to those visually impaired people who live in small towns and villages and thus cannot enjoy cinema screenings or theatre performances with AD, as these are usually organised in large cities only. Furthermore, TTS AD allows spectators with visual impairments to watch films and other audiovisual programmes on their own, without depending on others or being restricted to the explanations of their sighted friends or family.[5]

Text-to-speech audio description can be used for both domestic and dubbed productions (where only one language is heard) and for foreign programmes (where two languages can he heard: the original and the translation). For foreign materials, it can be combined with either audio subtitling of the dialogue (in subtitling countries, cf. Orero and Braun 2010) or with the voice-over translation (in Poland).

As with anything, text-to-speech AD does have its downsides. First of all, it requires media literacy and, as such, it largely excludes visually impaired people, especially the elderly, who live outside modern high-tech information society and do not interact with the new digital technologies. Another important criticism directed at TTS AD which can be anticipated is rooted in the fact that it does not serve to promote integration and inclusion as the viewing takes place at home, and often only involves one visually impaired person. In response to this anticipated criticism, I can only say that the arrival of the DVD/home video has not sent cinema to the dustbin of cinematographic history. In the same way, the aim of text-to-speech AD is to complement, not eradicate, the experience of watching films. TTS AD is by no means intended to replace the audio description practice currently in use. Rather, it aims to supplement it and to increase the number of audio described films and audiovisual programmes made available to people with visual impairments.

Previous studies on synthetic speech

In their everyday lives, visually impaired people can benefit from numerous text-to-speech applications, ranging from "leisure activities to devices which

support independent living" (Cryer and Home 2008: 5). Among them are various mobility aids, educational tools, screen reading software and entertainment (Freitas and Kouroupetroglou 2008, in Cryer and Home *ibid*.).

Previous studies on the acceptance of synthetic speech by visually impaired people focused on, *inter alia*, reading daily newspapers (Hjelmquist *et al*. 1990), receiving financial information (Thompson *et al*. 1999) and listening to a GPS system (Loomis *et al*. 2005, all in Cryer and Home 2008). It has been found that while synthetic speech may be difficult to comprehend at first, its tolerance and comprehension increases with more exposure and experience. In addition, while most people would prefer a natural voice, they find synthetic voices acceptable for a number of applications.

A study conducted by the RNIB on the attitude of blind and partially sighted people towards the application of synthetic speech for the Talking Books service has revealed that most people prefer a human narrator over a synthetic voice when it comes to reading books (Cryer and Home 2009). It has also been found that "most users felt synthetic speech would be acceptable for reference, instructional and non-fiction books, whilst fiction and leisure reading would be preferable with a human narrator" (Cryer and Home 2009: 5-6). It would be interesting to research whether similar patterns can be observed among users with regard to fiction and non-fiction audiovisual programmes with human vs. text-to-speech audio describers.[6]

The perception of text-to-speech audio description by visually impaired viewers

What follows is an overview of the first part of a three-stage study aimed at investigating the feasibility of using text-to-speech audio description in fiction and non-fiction audiovisual programmes. The three stages to be pursued in the research include the implementation of TTS AD in:

1. a Polish feature film,
2. a foreign feature film in English combined with voice-over in Polish,
3. a documentary in English combined with voice-over in Polish.

Rationale

The feature film selected for the first stage of the experiment was *Dzień Świra* (*The Day of the Wacko,* 2002, dir. Marek Koterski), a tragi-comedy telling the story of a middle-aged Polish literature teacher, Adam Miauczyński. Here is a brief summary of the plot, published in *The New York Times*:

> Writer/director Marek Koterski's dark comedy "Day of the Wacko" follows Adam over the course of a long, typically unpleasant day as he deals with his noisy neighbors, his overbearing mother, his apathetic son, his bitchy ex-wife, his rudely flatulent students, and, most debilitating of all, his own obsessive-compulsive behavior, and his immobilizing despair over the state of his life and the world around him. All the while, he reminisces about the woman he calls his great lost love, Ela, and fantasizes about seeing her again. Reaching a fever pitch of depressed paranoia, Adam decides to travel to take a train to the beach to find some peace. After a harrowing trip, during which he's forced to share a compartment with a motley assortment of obnoxious fools, he arrives at the sea and lies out in the sand, hoping for a moment's tranquillity as he continues his ongoing internal monologue, analyzing the failures of his life and his world. Ralske (2002: online)

The film seems to lend itself quite well to AD, albeit not without problems. A large part of the film narrative is Adam's voice-over commentary, his inner voice reporting his thoughts and feelings to spectators. In the exposition phase of the film, i.e. "the portion of the plot that lays out important story events and character traits in the opening situation" (Bordwell and Thompson 2008: 86), Adam narrates his every-day routine, with the camera showing him simultaneously performing the actions he is describing. The main character thus turns into a self-reflexive audio describer. The narration in *The Day of the Wacko* is restricted to Adam's vantage point. The plot plunges into his mind and spectators are presented with the events filtered through his perception. It is through Adam, the character-narrator that we learn what he thinks and feels in almost every scene.

It was decided that the AD script would be read by a female voice. This choice was motivated by three major factors. The first one stemmed from the fact that *The Day of the Wacko* consists largely of the main protagonist's monologues, interspersed with him conversing with, or rather barking at, other characters. Since the main character is male, it was thought that it would be easier and less confusing for viewers to listen to audio description delivered by a female voice.

The second factor contributing to the selection of the female voice for synthetic AD was the unquestioned hegemony which has so far been enjoyed by male voice talents in Poland, where the study is carried out. Poland is a country where the dominant mode of audiovisual translation on television is voice-over (Szarkowska 2009), both for fiction and non-fiction genres. The overwhelming majority of voice-over artists, known as *lektors*, are male. With the advent of pre-recorded audio description in Poland, it was only natural for many people, accustomed to hearing male narrators, and for the lobby of *lektors* themselves, that AD should also be read by men. Hence, on all the DVDs with pre-recorded AD released on the Polish market so far, as well as on tens of hours of audio described TV series produced by public television (TVP) and made available online, the audio describer is always male.

Finally, as the text-to-speech project develops and reaches stage 2, i.e. the synthetic AD of a foreign film, where by definition there will be a male voice reading the translation of all the characters' utterances, it was decided that a female voice for AD is necessary to facilitate comprehension. The processing effort expended by visually impaired viewers when watching a film—where on top of a number of original actors' voices there will be the male voice-over artist's voice—was deemed challenging enough not to be compounded by another male voice.

Problems related to audio description

Essentially, the difficulties related to audio describing *The Day of the Wacko* were twofold: (1) those stemming from the specificity of text-to-speech AD and (2) those typical of any audio description, regardless of the medium. Among the problems relating to the creation of TTS AD were foreign words and abbreviations. At the very beginning of the film, there are logos and names of three film companies involved in the production. It so happened that two out of the three mentioned names were English: Vision Film Distribution Company and Non Stop Film Service (see Fig. 2).

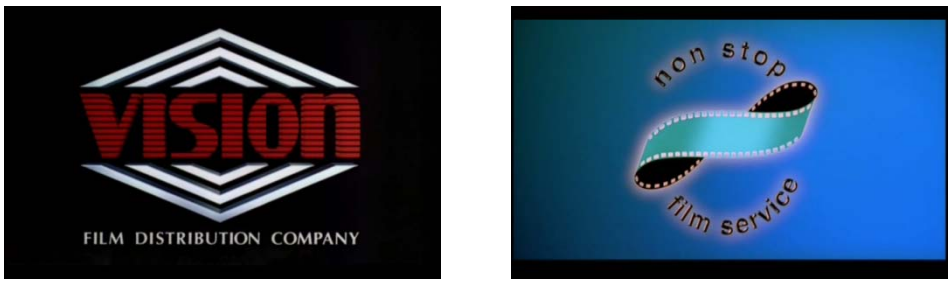

Fig. 2 Screenshots from the beginning of *The Day of the Wacko*

As the language of text-to-speech software was set to Polish, the AD script was read out according to Polish pronunciation rules, which turned out to be unacceptable. To overcome this problem, it was necessary to transcribe the names, approximating their pronunciation as much as possible to natural English, the end result being, respectively: "viżyn film distribjuszn kampany" and "non-stop film serwis." A similar strategy had to be adopted for dealing with abbreviations, such as EKG (electrocardiogram), which had to be re-written in the TTS AD script as 'e-ka-gie' so that the software could read it in natural Polish.

Another problem in TTS AD was related to stating the time. In one scene, having declared that he likes to start doing things exactly on the hour or at worst at half past the hour, Adam snuggles down on a sofa and looks at the clock, which shows 8:33 am. Disgruntled, Adam gets up and resets the clock to 8:29, then waits for 8:30 sharp and begins to read. In Polish, the time is given

using ordinal numbers, for example in 8:33, the '8' becomes 'ósma' ['eighth'], not the nominative 'osiem' ['eight']. The TTS software, however, reads out the numbers in the nominative, here as 'osiem' and not as 'ósma.' Consequently, all the times in the AD script had to be changed from numbers (8:33) into properly inflected full words ('ósma trzydzieści trzy' ['eight thirty three']. Another option was to change TTS software dictionary configurations and add the hours as exceptions, defining how we want the software to read particular numbers.

The greatest downside of text-to-speech AD at this stage of development seems to be the unnatural intonation and incorrect pronunciation of some clusters. The speech synthesis software used in the experiment revealed a few minor lapses, for instance in short sentences ending in the reflexive particle *się*, such as *Wierci się* [He is snuggling down] or *Zaciąga się* [He's dragging on a cigarette], the stress fell on the particle instead of the verb, producing an effect which was far from natural.[7]

Related to intonation is the question of punctuation, as the latter can have a significant impact on the former. To illustrate this, let us quote two famous contradictory readings of the phrase: (1) 'Woman without her man is a savage' and (2) 'Woman: without her, man is a savage.' In view of the above, whenever a short pause was required, a comma had to be inserted, while longer pauses required the use of a full stop.

Last but not least with regard to TTS AD, a major issue to be worked on in the near future to improve its quality is the correlation of the AD reading speed with the length of each AD chunk ('AD subtitle'). In the course of the experiment it has turned out that at present the subtitle reader only takes into consideration the in-time, i.e. the beginning of each AD chunk. It reads the AD script based on the reading speed set by the user. This means that the end of each AD chunk is not at present correlated with the out-time set in the subtitling software. I believe this problem is a minor technicality which can be easily solved, so that the speech synthesis software could read out the script at different paces, depending on how much time there is available for the AD.

Among the general difficulties of audio describing *The Day of the Wacko* was the lack of time to describe what was being shown on screen. This was due to the co-existence of dialogues and Adam's commentary, which often left no gap for audio description and was particularly true in the case of a series of rapidly intercut shots displayed simultaneously with Adam's interior monologue.

Another typical AD problem was the moment of naming the main character. *The Day of the Wacko* has become a classic since its release in 2002; consequently, the name of the actor playing the main character (Marek

Kondrat) and the character's name in the film (Adam Miauczyński) are widely known in Poland. The name of the main character, however, appears for the first time as late as in minute 47 of the film. It is during his acupuncture appointment that Adam is asked about his name and age, to which he promptly replies: "Miauczyński Adaś. Adam."[8] In spite of this delayed introduction, it was decided that Adam's name should be given in the AD script from the outset.

It is generally believed that the language of the AD script should suit the language of the film (ADI, n.d.). The "match your vocabulary to the show" injunction (*ibid*.) appears to be somewhat controversial in the case of films such as *Trainspotting* (dir. Danny Boyle, 1996), which contains strong language. Moreover, the rule seems to be based on the assumption that the language used in the entire film is uniform in terms of register and style. This, in fact, is not the case with *The Day of the Wacko*. In the film, Adam strives to speak 'proper' Polish, often correcting himself and pronouncing certain grammatical endings hypercorrectly. At the same time, however, he curses and swears both in his interior monologues and when talking to other characters. The strong language serves to emphasise his emotions, but would be uncalled-for in the AD script, particularly one read by speech synthesiser.

After the preparation of the TTS AD script for *The Day of the Wacko*, it was time for it to be tested with visually impaired viewers. The script was first discussed with two partially sighted people, whose remarks were taken into consideration in producing the final version of the script.

Research questions

The key objective of the present study was to determine whether visually impaired viewers would find it acceptable for text-to-speech software to read AD scripts. To address this objective, the following three research questions were formulated:

1. Which AD voice would the visually impaired prefer if they had a choice between a human voice and a synthetic voice?

2. Would TTS AD be acceptable as an interim solution, until there is more AD available with human narrators?

3. Would TTS AD be acceptable as a permanent solution, next to AD read by a human voice?

Procedure

The questionnaire[9] was administered after a screening of *The Day of the Wacko* with text-to-speech AD on 4 December 2009. The screening was part of the conference *Reha for the Blind in Poland*, which took place in Warsaw and saw the participation of numerous members of the blind and partially sighted community. It was the first screening of this kind in Poland.
The audience was first invited to watch the film and after the projection they were asked to provide answers to 15 questions, which were read out to them by myself and several other sighted volunteers. The first part of the questionnaire was meant to establish the participants' age, education, degree of sight loss (mild, moderate, severe, profound) and type (congenital/acquired). The second part aimed to find out their views on the use of speech synthesis in AD. They were also asked about their previous experience with audio description as well as their familiarity with computers, the Internet and speech synthesis software.

Sample

A total of 24 people were interviewed (13 females, 11 males). Five were aged 18-25 years (3 females, 2 males), ten were aged 26-39 years (5 females, 5 males) and nine were aged 40-59 years (5 females, 4 males).
As for educational background, four respondents had primary education, eleven—secondary, and nine were university graduates.

Out of the total number of participants (n=24), sixteen (66%) were congenitally blind and six (33%) had an acquired sight loss. The level of participants' sight loss was classified into four categories: mild, moderate, severe and profound. The scale used in the questionnaire was adopted from the research conducted by the RNIB (Freeman *et al*. 2008), which was based on the Network 1000 research report (Douglas *et al*. 2006).[10]

As seen in Figure 3, eight respondents (33%) had profound sight loss, three of them (13%) had severe sight loss, eight had moderate sight loss and five of them (21%) had mild sight loss:

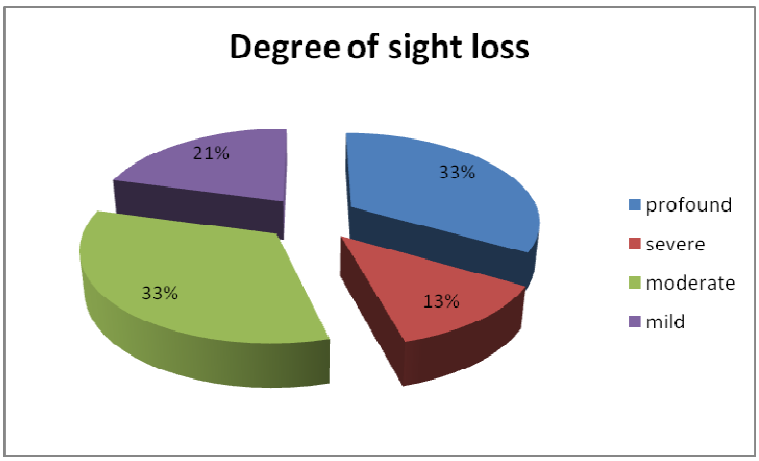

Fig. 3 Degree of sight loss among respondents (n=24)

Results

With respect to research question (1), i.e. the preference for either a human or a synthetic voice to read out the AD script, the majority of respondents (n=13, 54%) stated they would prefer a human voice whilst two people (8%) claimed they preferred a synthetic voice over a human one (Fig. 4). As many as one in four declared that the choice of human vs. synthetic voice depended on the programme. Many others were not sure and wanted to have more experience with TTS AD in order to be able to make an informed choice.

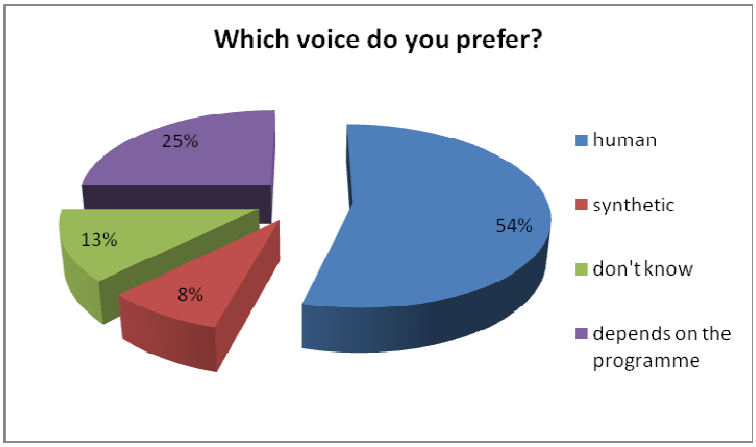

Fig. 4 Reseach question 1 (n=24)

No significant patterns were found in terms of the preference for human/synthetic voice depending on the age, gender, education of the respondents or their previous exposure to synthetic voice. In order to formulate further generalisations on each of these variables, a larger sample would need to be examined.

Research questions (2) and (3) addressed the acceptance of TTS AD as either

an interim or permanent solution (Fig. 5). 95% of respondents (23 out of 24) were in favour of introducing TTS AD as an interim solution until there are more programmes available with human audio describers. Almost two in three respondents (n=14, 58%) supported TTS AD as a permanent solution, functioning next to AD with a human voice. One third (n=7, 29%) were against TTS AD as a permanent solution. Some respondents (n=3, 13%) said they would need more time and experience with TTS AD to make an informed choice.

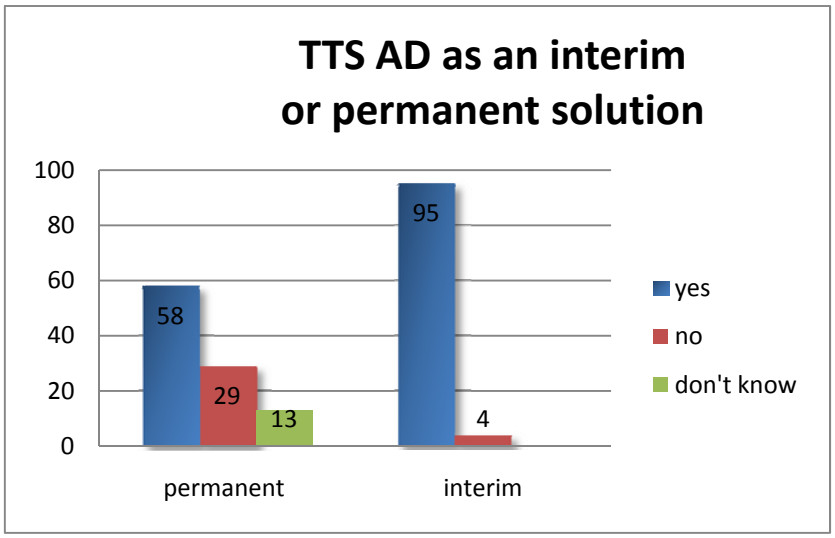

Fig. 5 Research questions (2) and (3) (in %, out of the total n=24)

The preference for TTS AD as an interim or permanent solution was then examined by different variables (Fig. 6), such as the type and degree of sight loss as well as the age of the respondents:

	TTS AD as an interim solution			**TTS AD as a permanent solution**		
	yes	**No**	**don't know**	**yes**	**no**	**don't know**
By type of sight loss						
congenital	94%	6%	0	47%	35%	18%
acquired	100%	0	0	85%	15%	0
By degree of sight loss						
mild	100%	0	0	80%	20%	0
moderate	100%	0	0	37%	25%	37%
severe	100%	0	0	66%	33%	0
profound	87%	13%	0	63%	37%	0
By age						
18-25	100%	0	0	40%	40%	20%
26-39	90%	10%	0	60%	20%	20%
40-59	100%	0	0	66%	33%	0

Fig. 6 TTS AD as an interim or permanent solution (n=24)

No significant patterns emerged in terms of the attitude towards TTS AD as an interim solution. As for TTS AD as a permanent solution, more respondents with acquired sight loss than those with congenital sight loss declared their support for the idea. Furthermore, perhaps somewhat surprisingly, there were more respondents from the older age brackets who declared their support for permanent TTS AD than respondents from the youngest group.

The familiarity of the respondents with computers, the Internet and speech synthesis software was found to be quite high. 21 out of 24 (87%) respondents have either a PC or a laptop at home and 18 respondents (75%) also have an Internet connection. The overwhelming majority of respondents (75%) use speech synthesis software on a regular basis, but only 5 people watch films with subtitles read out by text-to-speech software (many respondents were surprised to hear this is possible and were willing to try it out).

When asked if they would like to choose their own synthetic voice for the reading of TTS AD, 23 respondents (95%) said "yes", and only one was not sure. This was confirmed by some participants who said they would prefer a different synthetic voice than the one used in the screening and suggested one of the voices they use most often. Other people stated that they understood why a female voice had been chosen, but declared that they would still prefer a male voice, as they are accustomed to listening to male voice talents.

As for previous exposure to AD, 14 respondents (58%) stated they had seen audio described films before, while for the remaining 10 (42%), the TTS AD was the first screening with AD they had ever experienced.

Discussion

The overall results of the present study are in line with the RNIB report on the use of synthetic speech by blind and partially sighted people, which states that "listeners prefer natural sounding speech, both in comparing natural speech to synthetic speech and in comparing different synthetic voices" (Cryer and Home 2008: 7). It is worth noting, however, that while the visually impaired viewers in this study find natural speech *preferable*, many of them would find synthetic speech *acceptable*.

The high level of familiarity with technology demonstrated by the respondents may not be representative of an overall visually impaired population. While TTS AD may not be a feasible solution for older age groups, whose unfamiliarity with computers, the Internet and speech synthesis software remains a serious obstacle, the results of the present research still demonstrate an untapped potential for text-to-speech audio description.

Apart from previous experience, another important factor affecting comprehension of speech synthesis is the rate at which the text is read. Several respondents complained that the rate set in the experiment was too high and therefore difficult to understand. This finding was used in further stages of our research and the reading speech parameter will be adjusted accordingly.

According to previous research, "synthetic speech suffers more degradation in the presence of background noise than does natural speech" (Papadopoulos *et al*. 2009: 405), which ends up affecting intelligibility and comprehension rates. This is probably due to the fact that human speakers adjust the volume of their speech to the level of surrounding noises, raising their voice when the background noise volume increases (Langer and Black 2005). In the present study, a few respondents complained that it was difficult to understand the TTS AD script when it was read over other sounds in the film, both diegetic and non-diegetic. If TTS AD were to be implemented on a larger scale, this shortcoming would need to be addressed.

All in all, while TTS AD was not found to be the *preferred* solution for the reading out of the AD script, the study has shown that many respondents find it *acceptable* either as an interim (95%) or an alternative (58%) solution to traditional AD, especially if it means that more audiovisual programmes will be available to the visually impaired.

Towards the future

Text-to-speech audio description opens up new research avenues under the umbrella of accessibility and translation studies. More research is necessary to find out whether TTS AD is more suitable for certain genres, be it fiction or non-fiction. This study has focused on text-to-speech audio description for a domestic (i.e. untranslated) feature film, but it would be interesting to see the results of applying TTS AD to foreign feature films as well as non-fiction programmes, such as documentaries or educational materials. Another research idea would be to investigate the feasibility of translating TTS AD scripts and producing AD templates. The question of copyright also needs to be addressed.

TTS AD may not be a good-for-all solution—it still remains to be seen whether it can be accepted by people in the older age groups, with acquired (as opposed to congenital) visual impairments and who are unaccustomed to hearing synthetic speech and/or also have hearing impairments. All in all, however, text-to-speech audio description seems to allow for faster and cheaper creation of audio descriptions, easier distribution of AD scripts and therefore a wider availability of audio described products. There is still a lot to be done in the area

of making audiovisual programmes accessible to the blind and visually impaired and TTS AD demonstrates considerable scope for further development.

Acknowledgements

Many thanks to Piotr Wasylczyk, Anna Jankowska, Robert Więckowski and Mateusz Ciborowski for their help with the drafting of the AD script; to Leen Petré from RNIB for her invaluable help and feedback on the design and interpretation of the questionnaire; to Ewa Nowik-Dziewicka for her perceptive comments on a draft version of this paper; to Marek Kalbarczyk from the Foundation of the Chance for the Blind for allowing me to organise the screening during the conference *Reha for the Blind*; to Ivo Software for letting me use the Ivona synthesiser at the screening; to all the friends and students who helped me distribute the questionnaire, and finally to all the respondents for participating in the experiment and sharing their opinions on TTS AD.

This work has been supported by research grant N° N N104 148038 of the Polish Ministry of Science and Higher Education for the years 2010-2011.

References

- **ADI (n.d.) "**AD Guidelines" www.adinternational.org
- **Benecke, Bernd** (2004). "Audio description." *Meta* 49 (1), 78-80.
- **—** (2007). "Audio Description: Phenomena of Information Sequencing." MuTra 2007 – LSP Translation Scenarios: Conference Proceedings. Online at: http://www.euroconferences.info/proceedings/2007_Proceedings/2007_Benecke_Bernd.pdf (consulted 10.12.2010)
- **Bordwell, David** and Kristin Thompson (2008). *Film Art. An Introduction*. 8th edition. New York: McGraw-Hill International Edition.
- **Braun, Sabine** (2007). "Audio Description from a discourse perspective: a socially relevant framework for research and training." Languages and Translation Papers from the Centre for Translation Studies, University of Surrey. Onlineat: http://epubs.surrey.ac.uk/cgi/viewcontent.cgi?article=1000&context=translation (consulted 10.12.2010)
- **Braun, Sabine** and Pilar Orero (2010). "Audio description with audio subtitling — An emergent modality od audiovisual localisation." *Perspectives* 18: 3, 173-188.
- **Chmiel, Agnieszka** and Iwona Mazur (2008). "Percepcja filmu a ogólnoeuropejskie standardy audiodeskrypcji – polski wkład w projekt Pear Tree" ['Film perception vs. European-wide audio description standards – Polish contribution to the Pear Tree project'].

Przekładaniec no. 20, Kraków: Tertium, 138-158.

- **Ciborowski, Mateusz** (2008). "Znaczenie audiodeskrypcji dla niewidomych w Polsce." ['The importance of audio description to visually impaired people in Poland'] *Przekładaniec* no. 20, Kraków: Tertium, 136-137.

- **Cryer, Heather** and Sarah Home (2008). "Exploring the use of synthetic speech by blind and partially sighted people." RNIB Centre for Accessible Information, Birmingham: Literature review #2.

- **—** (2009). "User attitudes towards synthetic speech for Talking Books." RNIB Centre for Accessible Information, Birmingham: Research report #7.

- **Díaz Cintas, Jorge** (2005). "Accessibility for All." *Translating Today* 4, 3-5.

- **Douglas Graeme,** Christine Corcoran, and Sue Pavey (2006) *Network 1000. Opinions and Circumstances of Visually Impaired People in Great Britain: Report Based on over 1,000 Interviews*. University of Birmingham, Visual Impairment Centre for Teaching and Research, School of Education.

- **Fels, Deborah,** John Patrick Udo, P. Ting, Jonas E. Diamond, Jeremy I. Diamond (2005). "Odd Job Jack Described – A first person narrative approach to described video." *Journal of Universal Access in the Information Society* 5(1), 73-81.

- **Freeman, Jonathan,** Jane Lessiter and Eva Ferrari (2008). "Research report: Are you really listening? The equipment needs of blind and partially sighted consumers for accessible and usable digital radio." London: RNIB.

- **Freitas, Diamantino** and Georgios Kouroupetroglou (2008). "Speech technologies for blind and low vision persons." *Technology and Disability* 20, 135-156.

- **Greening, Joan** and Deborah Rolph (2007). "Accessibility: raising awareness of audio description in the UK." In: Jorge Díaz Cintas, Pilar Orero and Aline Remael (eds) *Media for All. Subtitling for the Deaf, Audio Description and Sign Language*. Amsterdam and New York: Rodopi, 127-138.

- **Hjelmquist, Erland,** Bengt Jansson and Gunilla Torell (1990). "Computer-oriented technology for blind readers." *Journal of Visual Impairment and Blindness* 17, 210-215.

- **Holland, Andrew** (2009). "Audio Description in the theatre and the visual arts: Images into words." In: Jorge Díaz Cintas and Gunilla Anderman (eds) *Audiovisual Translation. Language Transfer on Screen*. Basingstroke: Palgrave Macmillan, 170-185.

- **Langer Brian** and Alan Black (2005. "Improving the understandability of speech synthesis by modelling speech in noise." *Proceedings of the 2005 International Conference on Acoustics, Speech and Signal Processing* at Pennsylvania Convention Centre, 265-268.

- **Loomis, Jack M., J**ames R. Marston, Reginald G. Golledge and Roberta L. Klatzky (2005) "Personal guidance system for people with visual impairment: a comparison of spatial displays for route guidance." *Journal of Visual Impairment and Blindness* 99 (4), 219-232.

- **Matamala, Anna** and Pilar Orero (2007). "Accessible opera in Catalan: Opera for all." In: Jorge Díaz Cintas, Pilar Orero and Aline Remael (eds) *Media for All. Subtitling for the Deaf,*

Audio Description and Sign Language. Amsterdam and New York: Rodopi, 201-214.

- **Ofcom** (2000). "ITC guidance for standards in audio description." Online at: http://www.ofcom.org.uk/static/archive/itc/itc_publications/codes_guidance/audio_description/index.asp.html (consulted 10.12.2010)

- **Orero, Pilar** (2007). "Sampling audio description in Europe." Jorge Díaz Cintas, Pilar Orero and Aline Remael (eds) *Media for All. Subtitling for the Deaf, Audio Description and Sign Language*. Amsterdam and New York: Rodopi, 111-125.

- **Orero, Pilar** and Sabine Braun (forthcoming). "Audio Description with Audio Subtitling – an emergent modality of audiovisual localisation." Accepted for publication in *Perspectives*.

- **Papadopoulos, Konstantinos** et al. (2009) "Perception of Synthetic and Natural Speech by Adults with Visual Impairments." *Journal of Visual Impairment and Blindness* 103 (7), 403-414.

- **Ralske, Josh** (2002). "The Day of the Wacko." *New York Times*. Online at: http://movies.nytimes.com/movie/284626/Day-of-the-Wacko/overview (consulted 10.12.2010)

- **Remael, Aline** and Gert Vercauteren (2007). "Audio describing the exposition phase of films. Teaching students what to choose." *Trans. Revista De Traductología* 11, 73-93.

- **Salway, Andrew** (2007). "A corpus-based analysis of audio description." In: Jorge Díaz Cintas, Pilar Orero and Aline Remael (eds) *Media for All. Subtitling for the Deaf, Audio Description and Sign Language*. Amsterdam and New York: Rodopi, 151-174.

- **Schmeidler, Emilier** and Corinne Kirchner (2001). "Adding Audio Description: Does It Make a Difference?" *Journal of Visual Impairment and Blindness* 95 (4), 197-212.

- **Snyder, Joel** (2008). "Audio description: the visual made verbal." Jorge Díaz Cintas (ed.) *The Didactics of Audiovisual Translation*. Amsterdam /Philadelphia: John Benjamins Publishing Company, 191-198.

- **Szarkowska, Agnieszka** (2009). "The audiovisual landscape in Poland at the dawn of the 21st century." Angelika Goldstein and Biljana Golubović (eds) *Foreign Language Movies – Dubbing vs. Subtitling.* Hamburg: Verlag Dr. Kovač, 185-201.

- **Szarkowska, Agnieszka** and Anna Jankowska (forthcoming). "Text-to-speech audio description of voiced-over films. A case study of audio described *Volver* in Polish."

- **Theunisz, Mildred** (2002) "Audio subtitling: A new service in Netherlands making subtitling programmes accessible." Available at: www.sb-belang.nl

- **Thompson, Leanne,** Chris Reeves and Kate Masters (1999) "In the balance: making financial information accessible." *British Journal of Visual Impairment* 17 (2), 65-70.

- **Udo, John Patrick** and Deborah Fels (2009a). "Re-fashioning fashion: an exploratory study of a live audio described fashion show." *Universal Access in the Information Society* 8 (3), 123-238.

- **—** (2009b). "'Suit the Action to the Word, the Word to the Action': An Unconventional Approach to Describing Shakespeare's Hamlet." *Journal of Visual Impairment and Blindness*

103 (3), 178-183.

- **Vercauteren, Gert** (2007). "Towards a European guideline for audio description." Jorge Díaz Cintas, Pilar Orero and Aline Remael (eds) *Media for All. Subtitling for the Deaf, Audio Description and Sign Language*. Amsterdam and New York: Rodopi, 139-149.

Biography

Agnieszka Szarkowska, PhD, is Assistant Professor at the Institute of Applied Linguistics at the University of Warsaw. Her research interests include audiovisual translation, especially audio description and subtitling for the deaf and the hard of hearing. Her current research projects include text-to-speech audio description, subtitling for the deaf and hard of hearing in multilingual films, and the application of eye tracking to subtitling. She is also a member of ESIST and honorary member of the Polish Audiovisual Translators Association (STAW).

Contact: a.szarkowska@uw.edu.pl

[1] A patent application was filed at the Polish Patent Office with regard to text-to-speech audio description (application no.: P-389342).

[2] In the US and Canada audio description is often referred to as 'video description'.

[3] Theunisz (2002) describes an experiment using speech synthesis software to read the subtitles in foreign-language programmes on Dutch television for visually impaired viewers (known as 'audio subtitling'). The experiment revealed that the participants were "very enthusiastic about this initiative", but they complained about "the quality of the sound of audio subtitles" and found the synthetic voice "monotonous, dull and without emotion" (*ibid*.). In a similar vein, the *ITC Guidance on Standards for Audio Description* (2000: 28) states: "Equipment is available which can read teletext-delivered subtitles aloud, but the

expressionless quality of a synthesised voice is not suitable for an entire drama or a film, and it is not feasible to recognise a variety of different speakers within the programme". However, the quality of text-to-speech systems has improved significantly over recent years, calling for more research in this area.

[4] Polish public television Telewizja Polska (TVP) has made a number of TV series with AD available on its website: www.tvp.pl. Viewers can download them for free after having obtained a password from the Polish Association of the Blind (PZN) or by purchasing individual episodes.

[5] This is hardly possible with the AD offered by TVP online, as there is no audio navigation and the website is not "friendly" to the visually impaired.

[6] Such work is presently carried out by myself (at the University of Warsaw) and by Anna Jankowska (Jagiellonian University.

[7] All instances of the unnatural pronunciation which occurred in the script were passed on to the TTS software manufacturing company, who promised to include the corrections in subsequent versions of the software.

[8] Adaś is a diminutive form of the name Adam.

[9] The questionnaire is reproduced here:

1. Age
   - □ 18-25
   - □ 26-39
   - □ 40-59
   - □ 60-74
   - □ 75+
2. Gender
   - □ Female
   - □ Male
3. Education
   - □ Primary
   - □ Secondary
   - □ Higher
4. Which of these best describes your sight with glasses or contact lenses if you normally use them but without any low vision aid? Imagine you are in a room with good lighting and answer yes, no or uncertain to each part please. Can you see well enough to:

□ Tell by the light where the windows are?	□ Yes	□ No	□ Uncertain
□ See the shapes of the furniture in the room?	□ Yes	□ No	□ Uncertain
□ Recognise a friend across a road?	□ Yes	□ No	□ Uncertain
□ Recognise a friend across a room?	□ Yes	□ No	□ Uncertain
□ Recognise a friend if he or she is at arm's length?	□ Yes	□ No	□ Uncertain
□ Recognise a friend if you get close to his or her face?	□ Yes	□ No	□ Uncertain
□ Read a newspaper headline?	□ Yes	□ No	□ Uncertain
□ Read a large print book?	□ Yes	□ No	□ Uncertain
□ Read ordinary newspaper print?	□ Yes	□ No	□ Uncertain

5. Type of sight loss

- □ Congenital
- □ Acquired

6. Do you use a PC or a laptop at home?
   - □ Yes
   - □ No
   - □ Don't know
7. Do you have an Internet connection at home?
   - □ Yes
   - □ No
   - □ Don't know
8. Do you use text-to-speech software regularly?
   - □ Yes
   - □ No
   - □ Don't know
9. If yes, do you watch films at home with subtitles read out by text-to-speech software?
   - □ Yes
   - □ No
   - □ Don't know
10. Have you seen any films with audio description?
    - □ Yes
    - □ No
    - □ Don't know
11. If you had a choice, which AD voice would you prefer?
    - □ A human voice
    - □ A synthetic voice
    - □ Depends on the film/programme
    - □ Don't know / doesn't matter
12. Would you accept TTS AD as an interim solution, until a system has been agreed to have a human voice reading out the AD?
    - □ Yes
    - □ No
    - □ Don't know
13. Would you accept TTS AD as a permanent solution, as an alternative to a human voice?
    - □ Yes
    - □ No
    - □ Don't know
14. Would you like to be able to choose a synthetic AD voice yourself, from your own selection of synthetic voices?
    - □ Yes
    - □ No
    - □ Don't know
15. Do you have any comments on the audio description to *The Day of the Wacko*?