The Routledge Handbook of Audio Description

Christopher Taylor, Elisa Perego

Artificial voices

Publication details


Agnieszka Walczak, Gonzalo Iturregui-Gallardo

Published online on: 07 Apr 2022

How to cite: Agnieszka Walczak, Gonzalo Iturregui-Gallardo. 07 Apr 2022, Artificial voices from: The Routledge Handbook of Audio Description Routledge

Accessed on: 22 Nov 2023


PLEASE SCROLL DOWN FOR DOCUMENT

Full terms and conditions of use: https://www.routledgehandbooks.com/legal-notices/terms

This Document PDF may be used for research, teaching and private study purposes. Any substantial or systematic reproductions, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The publisher shall not be liable for an loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.
1. Introduction

Embraced by the rapidly developing field of media accessibility, audio description (AD) “deals with sound track perception and seeks to convey and complement the message of the audiovisual text” (Oncins & Orero, 2020: 194). Two other accessibility services that make up the big umbrella of AD are audio subtitling (AST; Matamala in this volume) and audio introductions (AI; Romero-Fresco in this volume). The former, also referred to as spoken subtitles (Ljunglöf et al., 2012; Verboom et al., 2002), can be defined as “the aurally rendered and recorded version of the subtitles. . ., mixed with the original sound track” and “usually read, sometimes acted out, by one or more voice actors” (Reviers & Remael, 2015: 52). The latter are “introductory notes, show notes or program notes” (York, 2007), “spoken by a single voice or a combination of voices lasting between five and 15 minutes” (www.vocaleyes.co.uk). Serving as an essential resource for the blind and partially sighted (BPS) audience to help them understand the audiovisual content, they are often provided together. What they all have in common is that they are rendered orally, with either human or artificial voice. The latter is the focus of this chapter.

2. Artificial voices: overview

The voice, which is naturally produced by most human beings, originates in the pharynx through the vibration of vocal cords and the following modification of sound waves through the oral and nasal cavities. The idea of artificially recreating the human voice has been around for at least two hundred years, since Christian Kratzenstein, a Russian doctor, who in 1779, after studying the acoustic characteristics of a number of vowels, constructed various kinds of pipe-shaped instruments that imitated the sound quality of such vowels (Schroeder, 1993). Other experiments followed this first attempt to recreate the human voice, such as the invention of the accordion-like speaking object developed by Wolfgang von Kempelen some years later, or the organ-shaped instrument created by Willis in 1938 (Schroeder, 1993). The first electrical voice synthesisers did not appear until the 1960s based on previous similar ideas such as the VODER (Voice Operating Demonstrator) and VOCODER (Voice Coder) systems.
developed by Homer Dudley and Bell Laboratories, respectively (Flanagan & Rabiner, 1973). These semi-electrical inventions opened the door to the development of voice synthesisers and present-day technology.

Today, artificial voices are broadly applied in several contexts. When it comes to accessibility services for the BPS, these kinds of voices are a normality for those who use smartphones, laptops and other similar technological devices (Caruso, 2012; Cryer et al., 2010; Freitas & Kouroupetroglou, 2008). Speech synthesis technology is commonly used in “sat nav”, “tannoy” announcements, for audio books (Cryer & Home, 2009) and eBooks (RNIB, n.d.), in mobility aids, educational tools, communication (Cryer & Home, 2008) – even as a tool supporting translators in their jobs (Figiel, 2014).

The benefits of artificial speech include “speed of production, confidentiality and the potential for artificial speech to deliver information which was not otherwise available” (Cryer & Home, 2008: 4). Artificial speech grants more autonomy to its users and can improve their ability to work competitively with sighted persons (Listeri et al., 1993), for instance, it enables one to receive sensitive financial information in a comfortable manner (Thompson et al., 1999). Further advantages of artificial speech are linked to media production. If used for audiovisual content, it could increase the output of shows available with access services, which, when produced in a conventional way – require longer timelines and greater financial resources. Access services with text-to-speech (TTS) technology do not require studio recordings with voice talents reading out the scripts, thus minimising production costs. BPS people can make use of speech synthesisers they own, and only a text file would need to be supplied to them (Szarkowska, 2011). From this perspective, it would also mean greater independence, as BPS people would be able to watch audiovisual content by themselves, without the help of their sighted friends.

The downsides of using artificial speech are related to the way we consume audiovisual content. In order to access it, media literacy is required, which could potentially exclude certain groups of users, including the elderly, people who do not interact with modern technologies or those who do not have Internet access (Szarkowska, 2011). Artificial voices may also present some limitations in terms of quality, including pronunciation and intonation issues, especially in the context of accessing audiovisual content (Holsanova et al., 2015; Nielsen & Bothe, 2008). Lastly, having the services of AD, AST and/or AI delivered with artificial speech would mean the viewing takes place at home and thus social integration and inclusion could be sidelined.

3. Artificial voices in AD, AST & AI

In this section, we will focus on the application of artificial voices in the area of media entertainment and specifically in AD, AST and AI. All three services can be human-narrated, but they can also be delivered to audiences thanks to TTS technology. Firstly, we will analyse the technical aspects of creating AD, AST and AI with artificial voices, then we will comment on their delivery modes and finally we will discuss the reception studies carried out with BPS people that aimed to gather their experiences with the said service(s) delivered using speech synthesis software.

3.1 Creation process

The process of creating TTS AD was explained by Szarkowska and Jankowska (2012: 86):

first an AD script is written and then, using subtitling software, it is synchronised with the film. Put simply, synchronising the AD script with a film means preparing a list of
AD chunks in the form of subtitles, each consisting of text and time codes. From the technical point of view, the process differs from preparing traditional subtitles only by the fact that the text is to appear in the gaps between the dialogues and not simultaneously with them. Later on, the text file is read by speech synthesis software while the film is played on a multimedia player, or the AD script can be recorded and mixed with the multimedia file.

Apart from the technical aspects of TTS AD file preparation, linguistic aspects must also be borne in mind in the process. As typically the language of the TTS software is set to the native language of its user, it applies the pronunciation rules specific for this native language when reading out the AD script. Therefore, foreign words that appear throughout audiovisual productions, including the names of people, cities, signs or abbreviations, logos, may sound unnatural. This may be circumvented by transliterating foreign words into the user’s native language (Mączyńska, 2011; Tor-Carroggio, 2020; Szarkowska, 2011; Walczak & Szarkowska, 2012). For example, when delivering English names in a Catalan AD, their orthography would have to be altered in order to correspond with the original pronunciation: “Kate” would be written as “Queit” to suit the Catalan phonetic system (Iturregui-Gallardo, 2019). Other elements to be taken into account are stress and intonation. For instance, for Polish, “sentences ending in the reflexive particle ‘się’ . . . may need to be omitted, as otherwise when read by speech synthesis the stress falls on the particle instead of the verb, producing an odd or unusual effect” (Walczak, 2010: 67). When it comes to intonation, depending on the language, the speech synthesiser may not take into account question or exclamation marks, and therefore produce monotonous speech that does not convey the elements of anger, fear or surprise, even though they are emphasised in the AD script (Mączyńska, 2011).

The process of creating TTS AST is similar to that of TTS AD, except for the first step that consists in translating the dialogues of the audiovisual production from source to target language. Then, the text is cut into chunks to create subtitles. Finally, the time codes are allocated; “the time when a subtitle appears on the screen (in-time) and the time when a subtitle vanishes from the screen (out-time)” (Mączyńska, 2011: 45).

When combining TTS AD with TTS AST, certain aspects have to be considered. Firstly, it has to be remembered that three different soundtracks can be heard by a viewer: (1) original voices of actors (if any); (2) translation of original dialogues and (3) AD track. In such a scenario, a very important factor is the selection of synthetic voices for AD and AST. In order “not to confuse the viewers” and “facilitate comprehension” (Mączyńska, 2011: 44), the voices used for AD and AST should differ, especially in multilingual productions. The decision on the use of voices will, however, depend on the good practices in place in a given country, the nature of the audiovisual product and financial resources available. Among possible solutions are: using one voice to read both AD and AST (cf. Braun & Orero, 2010), applying two voices – for instance, male and female – one for each modality (cf. Mączyńska & Szarkowska, 2011), or employing a collection of voices: one for AD and a set for AST corresponding to each character (cf. Remael, 2012). If analysing possible voice combinations for AST solely, the options include:

one voice for all characters (as in voice-over), one voice per character (as in dubbing), one voice per gender, where the voices would be gender-matched, i.e., a female voice for all the female parts and a male voice for all the male parts.

(Szarkowska & Jankowska, 2015 198)
Another aspect of combining TTS AD with TTS AST is of a technical nature. The AD script for TTS software must be inserted in-between the dialogues translated for AST, and the in-time and out-time codes must be allocated. Also, voice tags for AD and AST must be added (Mączyńska & Szarkowska, 2011). The last step in the process is the synchronization of AD and AST tracks, which can also pose challenges. Those may be technical, such as setting the appropriate reading speed for the synchronization, or linguistic, such as reducing the text due to overlaps or adding connectors for character identification (Benecke, 2012; Braun & Orero, 2010; Mączyńska & Szarkowska, 2011). Also, depending on whether the “voice-over effect” or “dubbing effect” (Iturregui-Gallardo & Matamala, 2020) is desired to be achieved, AST may be delivered with a lag or in synchrony with the original dialogues (Braun & Orero, 2010).

The creation process of TTS AI is not as complex as TTS AD and TTS AST. The script is not prepared in the form of subtitles, but as a regular text and does not require synchronization, only setting up the reading speed, which can be adjusted by users depending on the needs. The major challenges, just as in the case of AD, may be associated with phonetic, pronunciation and grammar issues, as well as the question of intonation or selection of voice to read the AI script (Walczak & Szarkowska, 2012). Also, in order to maintain the flow and naturalness of speech, it is worth verifying the punctuation in the script. Some punctuation marks may need to be occasionally omitted or added, like a comma for short pauses or full stop for long pauses, irrespective of the grammar rules of a given language (Walczak, 2010; Szarkowska, 2011).

### 3.2 Delivery mode

The use of voice, be it artificial or natural, can be linked to the characteristics of the audiovisual content being adapted through AD, AST and/or AI. The main distinction is to be made between audiovisual products that are pre-recorded, namely screen audiovisual contents and those that are live, such as the scenic arts: theatre, opera and dance, even museums.

Pre-recorded AD, AST and/or AI are those services in which the sound contents made available to the users were either recorded, by means of a microphone, in the case of a human voice, or created, by means of TTS systems, in the case of an artificial voice. The files derived from these processes can be integrated in the main soundtrack of the audiovisual content or delivered live manually or automatically. When the files are integrated in the audiovisual content – usually a film, TV series, documentary or recorded scenic arts performance – the AD track can be introduced on top of the other tracks. With on demand services, the AD track can be added or removed depending on the wishes of the user. If AST is required, it is usually included as part of the AD track. However, AST can also be delivered as a separate track that may be activated or deactivated on demand. When AST is offered separately from the AD, screen readers normally read aloud subtitles using the artificial voices at the end-user’s receiver. Such autonomous AST is commonly found on television in countries with strong subtitling traditions, such as the Netherlands, Sweden, Denmark or Finland. It is made available to final users by means of systems such as TTS and Optical Character Recognition (Iturregui-Gallardo, 2019). As for AI, if its duration is short, it is usually built into the AD track at the beginning of the audiovisual product. For longer AIs, several solutions are available: they can be treated as separate files and added to the DVD/Blue-ray copy of the audiovisual product together with the AD track, stored in the cloud-based service of the application enabling access to audiovisual content for the BPS (Jankowska & Walczak, 2019), made available on the website or sent to BPS users by e-mail (www.vocaleyes.co.uk/).
Live AD, AST and/or AI are typically offered for non-screen artistic productions, however and due to the nature of the audiovisual product, they are mostly human-voiced. The example here can be the opera Roberto Devereux by Donizetti, one of the first user experiments on live AST in opera performances, where the AST delivered by a voice talent proved to be quite effective and accepted by the BPS audiences (Orero, 2007).

3.3 Reception studies

As for the media services provided under the umbrella of AD, reception studies with BPS people suggest that in general, human voices are preferred. They are easier to follow, especially in films where audiences are encouraged to experience changes in feeling, and particularly for children and young people (Holsanova et al., 2015). However, the research also shows that the acceptance of artificial speech may depend on the users’ experience, “as people were found to ‘get used to’ artificial voices” (Cryer & Home, 2008: 4; see also: Cryer & Home, 2009; Kobayashi, 2010; Szarkowska & Jankowska, 2012; Walczak & Szarkowska, 2012). Some people are willing to accept them as an interim solution or if it allows them to access information otherwise inaccessible (Cryer & Home, 2008). Among factors influencing the acceptance of artificial speech in audiovisual productions, BPS people mention the following: naturalness of the voice (Stevens et al., 2005; Ilves & Surakka, 2013), characteristics of the voice, like pleasantness or listening effort (Cryer & Home, 2008; Fernández-Torné & Matamala, 2015), prosodic features of the voice (Iglesias Fernández et al., 2015; Fryer & Freeman, 2014), genre (Ramos, 2015; Ramos Caro, 2016; Thrane, 2013) or narration speed (Cabeza-Cáceres, 2013).

Below is a summary of a selection of reception studies that focused on exploring BPS people’s experience with artificial voices for audiovisual productions. They are presented in sections devoted to AD, AST and AI respectively.

3.3.1 Audio description

Up to date, the reception of AD delivered with artificial voices has been tested in various contexts, including Japanese, American, Polish, Catalan, British and, more recently, Chinese.

The tests in Japan and the United States were conducted by Kobayashi et al. (2010). Their aim was to assess the acceptability and effectiveness of ADs delivered with an artificial voice. The study in Japan involved 115 participants with vision impairment who were asked to watch a fragment of a movie and a recipe clip, with human-narrated and synthesised AD. The findings show that ADs delivered with a human voice were ranked best by most participants, with TTS being the second favourite. For a quarter of participants though, the TTS voice was a preferred option. One of the reasons for that was because “the synthesised narrations allowed the participants to clearly distinguish the video descriptions from the original audio tracks” (Kobayashi et al., 2010: 164). The study in the United States, on the other hand, involved 236 BPS people who watched five clips of a different nature: informational, educational and entertainment. Although most participants preferred human-narrated ADs, the majority found “listening to TTS descriptions to be a comfortable, acceptable, or neutral experience” (Kobayashi et al., 2010: 165). Overall, both studies point to the conclusion that ADs with artificial voices are generally accepted by non-European BPS users, irrespective of the language, but they also highlight that TTS ADs may be more suitable for informational content, whereas human-narrated ADs for entertainment.

A year later studies began in Poland. The first one, carried out by Szarkowska (2011), was aimed at testing TTS AD of a monolingual feature film “Dzień Świra” [“Day of the Wacko”]
(dir. Marek Koterski, 2002). 24 BPS people were interviewed after the screening of the film. When asked for a preferred voice for AD delivery, the majority opted for the human voice. However, all but one were for introducing TTS AD as an alternative to human-narrated ADs until a higher number of audiovisual products with AD are available on the market. Many respondents also supported the idea of having TTS AD as a permanent solution, functioning next to AD delivered by a human. Another study, conducted by Drożdż-Kubik (2011), concentrated on analysing the perception of TTS AD of a dubbed feature film “Harry Potter and the Philosopher’s Stone” (dir. Chris Columbus, 2002). Seventeen BPS teenagers took part in the screening listening to the audio described dubbed version of the film. Afterwards, they were interviewed by volunteers. The results of the study indicate that the majority of participants found TTS AD an acceptable, but not a preferred solution for AD delivery. In their study, Walczak and Szarkowska (2012) focused on implementing TTS AD to a dubbed educational television series for young audiences “Once Upon a Time . . . Life” (dir. Albert Barillé, 1987). A total of 76 BPS children took part in the test, aged between eight and seventeen years of age. The questionnaire was administered after the screening. Walczak and Szarkowska (2012) relied on comprehension questions, asking about the elements mentioned in the AD track. The majority of participants answered the questions correctly, reporting that they gained new information after watching the episode. The acceptability of applying TTS AD in educational content was also verified. Here, the responses were varied, with several negative opinions related to intelligibility and speed. Nevertheless, more than half of the participants expressed their enthusiasm about the TTS voice and were willing to watch other episodes of the series.

Another experiment followed shortly after. Szarkowska and Jankowska (2012) explored the application of TTS AD with voice-over to a foreign fiction film “Volver” (dir. Pedro Almodóvar, 2006). Twenty BPS participants took part in the study, answering a post-questionnaire after the screening. When asked about preferences, most of them opted for natural speech, but many were in favour of TTS AD as an interim solution, provided that it led to an increase in the availability of audiovisual products with AD.

Catalan BPS audiences also had a chance to experience TTS AD. Fernández-Torné and Matamala (2015) tested the feasibility of implementing TTS AD to dubbed feature films, as compared to human-voiced AD. 67 BPS people took part in the experiment. They assessed the voices based on several factors, linked to users’ reception, including overall impression, accentuation, pronunciation, speech pauses, intonation, naturalness, pleasantness, listening effort and acceptance. The results show that natural voices obtained statistically higher scores than artificial speech and are a preferred solution for voicing AD. However, users’ feedback on TTS AD was also positive, with 94% accepting synthesised AD as an alternative solution to AD delivered by a voice talent. When asked about the application of TTS AD to various content types, over 70% opted for documentaries, series and films.

Fryer and Freeman (2014) examined the application of TTS AD in the UK. Contrary to the previous studies, they focused on measuring the impact of AD delivered by a human and TTS software on emotion elicitation, empathy and presence in BPS people. Nineteen participants took part in the test, each of them watching six clips, one from each emotion category. Comparing TTS AD with human-narrated AD, only the latter positively enhanced presence and emotion elicitation. However, as the authors concluded, the perception of voices may also vary depending on the content type, with human-voice AD being more important for emotive content and synthesised AD for informative programmes.

Walczak and Fryer (2018) also measured the impact of AD on the emotional experience of BPS people. The novelty of their research lay in its methodology. They compared artificial and human-narrated ADs for emotive and non-emotive content directly with each other, not in
isolation. Also, their study aimed to check not only users’ opinions and preferences, but also the impact the AD voice has on BPS people’s emotions. To this end, presence measures were used. The results of their study, which involved 36 BPS people, show that human-narrated AD prompted significantly higher levels of presence for emotive content, with no statistically significant differences as to AD voice for non-emotive content. It is worth noting though that the scores for TTS AD and human-voiced AD were almost the same for non-emotive content. Also, in the post-questionnaire, 86% of participants expressed their willingness to watch non-emotive content with artificial speech and 19% were in favour of TTS AD for emotive content.

Recently, Tor-Carroggio (2020) studied TTS AD in the Chinese context. Firstly, she attempted to evaluate TTS AD in Chinese by comparing it with standard human-voiced AD regarding key features. Secondly, her aim was to verify whether TTS AD would be accepted in China as an interim solution until there is more AD available with human narrators, a permanent solution, or both. Forty participants took part in the study that was conducted using clips from a Chinese historical movie. They were asked to respond to a questionnaire that was designed to rate the voice in the clips presented. It contained questions related to overall comprehension and parameters evaluating the listening experience. The results show that the human voice scored statistically higher than the artificial voice. However, as many as 20% of participants liked the artificial voice better. It was “praised for being clear . . . very smooth and did not have an accent” (Tor-Carroggio, 2020: 183). Also, more than half of the participants were willing to accept it as an interim alternative and even as a permanent solution if this would lead to the increase in the amount of films with AD. The participants also considered the artificial voice to be more suitable when it comes to delivery of AD for non-fiction content. This, in turn, coincides with the findings of Kobayashi et al. (2010), Fernández-Torné and Matamala (2015) and Walczak and Fryer (2018).

3.3.2 Audio subtitling

The reception of AST delivered through artificial voices has been tested on only few occasions. Maćzyńska (2011) put to the test the AD to the non-fiction film “La Soufriere” with AST, with both tracks delivered through TTS technology. User reception was tested by means of a preference questionnaire that was administered to 84 BPS persons. The questionnaire was composed of twelve close-ended and three open questions. The results of the study showed that most participants found artificial voices for both AD and AST acceptable and suitable for this type of content. A vast majority of respondents thought AD and AST delivered with an artificial voice could be accepted as an interim solution (83%), at least until more products with AD/AST were made available. A large number of participants (69%) also considered that artificial voices could be an alternative for traditional human-voiced AD.

In Thrane (2013), the AST service created for the Danish channel DR1Syn was tested with sixteen participants, who were BPS, elderly, cognitive-impaired or learners of Danish. AST was tested in five-minute videos belonging to three different genres: news, documentary and fiction. Data was collected for each of the videos by means of a combination of methods: a click test, in which participants were asked to click on a button each time they experienced problems with AST, an interview and a self-report questionnaire on the quality of AST. The results showed that the main problems encountered by the participants were related to synchronisation, the use of multiple voices or just one voice, the fade level of the programme audio track and quality related to pronunciation and speed. These problems were more or less relevant according to the genre of the video. Finally, participants suggested that the AST service worked best for the news and least with fiction. The study concludes highlighting that
the acceptance of AST with artificial voices is linked to the users’ needs and the characteristics of the content.

In the experiment carried out by Holsanova et al. (2015) in Sweden, among the participants were only people with dyslexia and other reading impairments, no BPS users took part. They compared the use of human voices and artificial voices in the rendering of audio written subtitles in different contents: a short film, report feature from Gaza with interventions in foreign languages, a report feature of a telephone interview in which the sound quality is too poor to be understood and a report feature with text captions. To collect the data, preference questionnaires with rating scales were used. The results show that participants were eager to have the audio rendition of the written text. Even if artificial voices were slightly easier to hear, participants preferred human voices, particularly in films and TV series. For the news report, artificial voices were accepted. Participants identified the presence of several voices (original, AST and spoken captions) as a major inconvenience, as they could lead to comprehension issues.

There are other studies centred around AST, but they involved a human narrator. One of them – also one of the first experiments conducted with the AST – was a study testing AST at the opera (Orero, 2007). The service was very well received by participants, who had to rate their experience after the performance. Other experiments with human-voiced AST focused on emotional activation in users following a methodology in which self-report instruments and psychophysiology were combined (Iturregui-Gallardo & Matamala, 2020; Iturregui-Gallardo, 2019). The results show that content with AST induced emotions in BPS audiences, but, as no statistically significant differences were observed, it remains an area calling for more research.

### 3.3.3 Audio introductions

The research on AI is quite recent. This part of the AD service, even if used regularly in the creation of audio described materials, has not received a lot of attention from scholars. In recent years a series of studies have been published on the reception of AI for cinema, but with human voices. These studies took place in the UK (Romero-Fresco & Fryer, 2013) with the films “Slumdog Millionaire” (2008) and “Man on Wire” (2008); in Italy (Di Giovanni & Morettini, 2012; Di Giovanni, 2014) with “Slumdog Millionaire” (2008) and finally, in Poland (Jankowska, 2013) with “Man on Wire” (2008). The results were similar in all of them, pointing out that AIs were considered helpful by participants and enhanced the filmic experience.

Until the moment of writing this chapter, no studies on the reception of AIs delivered with artificial voices have been published. It is worth highlighting though that the use of artificial voices is a regular market practice, for instance, in the scenic arts (Hermosa-Ramírez, 2020). For this type of audiovisual content, the AD prepared by the audio describer is later on rendered in its audio form through artificial speech. Such AD also contains the explanation of the performance in the form of AI.

### 4. Conclusion

The aim of this chapter has been to account for the practice of using artificial speech in delivery of AD and its concomitant services, namely AST and AI. Audio description is “a complex process of cross-modal mediation with a plethora of challenges” (Braun & Orero, 2010: 176). When analysing AD, as pointed out by Fryer (2010: 205), “there is a tendency to focus on AD as a written text. A key element, easy to overlook, is that AD is always received aurally, together with the existing dialogue and soundscape”. Voice is a means that allows the audience
to become immersed in a story and enjoy it. Sound, just like vision, influences the reception of the audiovisual product and helps in creating “a relaxed experience” (Kobayashi, 2010: 165). Therefore, the choice of delivery voice should not be neglected.

The contrast between natural and synthetic voices for delivery of AD and AD-related services has been studied by various scholars in terms of acceptance and user preference. The studies indicate that natural voices are generally rated higher than synthetic voices. However, while the BPS audiences “find natural speech preferable, many of them would find synthetic speech acceptable” (Szarkowska, 2011: 155). They seem to be suitable, for instance, for documentary or instructional videos and for short content, where understanding is more crucial than pleasure (Kobayashi, 2010). Also, BPS people seem to be in favour of applying artificial voices to audiovisual products as an alternative solution, until more content accessible to the BPS becomes available. Certain gaps in knowledge of the reception of artificial voices are still to be addressed. Although for many Western languages natural voices are preferred, this is not the case, for instance, for Chinese. The explanation might be related to the linguistic differences of the languages. Chinese is characterised by its tone system, which might have implications in the way artificial voices are received and comprehended. Yet further research is required on the topic.

At present, many commercially used artificial voices are almost natural sounding and have a potential to be implemented on a wider scale. They are meant to complement, not to replace, current practices and make audiovisual content more readily available for the BPS. Furthermore, they could be of use for people in the need for easy-to-understand content since they can be seen as a means to neutralise accents or standardise speech (cf. EASIT EU Project, https://pagines.uab.cat/easit/en). With technological advances on the horizon, artificial voices may turn out to be a time and cost-effective alternative to traditional media access services provided by human voices.

5. References


Agnieszka Walczak et al.


7. **Online references**

https://pagines.uab.cat/easit/en
www.mib.org.uk/
www.vocaleyes.co.uk/