The Routledge Handbook of Audio Description

Christopher Taylor, Elisa Perego

Audio description software tools

Publication details
Vincenza Minutella
Published online on: 07 Apr 2022

How to cite :- Vincenza Minutella. 07 Apr 2022, Audio description software tools 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.
Audio description software tools

Vincenza Minutella

1. Introduction

Audio description (AD) can be created with the help of technology. Several AD tools exist which could be used to make the audio describing process faster and easier, or to train audio describers. The aim of this chapter is to provide an overview of the AD software tools available on the market or freely available online, to illustrate how such tools can be used to create audio descriptions and in the training of audio describers and to discuss what might be the pros and cons of these applications. The first part of the chapter will comment on the current situation regarding the use of AD software tools in academic and professional environments. It includes comments by practitioners and scholars from different countries on the use of such tools. Section three will describe some audio description software tools. Section four will report on an experiment carried out with a group of Master students at the University of Turin in 2015, while section five will focus on a new software called AD Author, the main features of which will be illustrated.

2. AD software in professional and academic environments

Neither AD trainers nor professional audio describers appear to make use of AD software to any great extent. Personal communications with academics and professionals, the ADLAB PRO Project (2016–2019) reports and recent studies seem to suggest that acquiring technology skills and teaching how to use specific AD software are not considered among the top priorities in AD training (ADLAB PRO, 2017a, 2017b; Mendoza & Matamala, 2019; Mazur & Chmiel, 2021). Mazur and Chmiel (2021) summarise the results of a survey carried out among 86 international AD trainers in approximately 192 academic and non-academic (vocational) courses. The survey was part of the ADLAB PRO project. The authors point out that “IT-related skills, such as using special software, was not prioritised in either of the two types of courses” and that “little prominence [is] given to technology in existing AD courses” (Mazur & Chmiel, 2021: 60). The scholars also highlight the high cost of “licensed professional software (such as Starfish Technologies and Miranda Softel Swift ADePT)”. This may be one of the reasons why such tools are not frequently used in academic courses.
the results of the survey carried out among trainers in Spain, Mendoza and Matamala (2019: 161) similarly comment that “Technical skills . . . are not covered in class to a large extent, the reasons being the lack of time or the fact that they are not considered necessary”. A further reason pointed out by Mazur and Chmiel (2021: 60) is that professionals themselves seldom use AD software:

ICT related skills received one of the lowest scores (3.51) of all the listed competences, irrespective of the type of course. In the case of non-academic settings and according to Bernd Benecke (personal communication), a professional audio describer employed by the Bayerischer Rundfunk in Germany, this might be due to the fact that working with specialist software is not very common in the industry, whether freelancing or being part of an in-house team.

Nevertheless, Mendoza and Matamala (2019: 161) argue that

even though these reasons might be in line with the professional profile demanded by the market in Spain, keeping up to date with future professional trends would be advisable so that students are ready for a global market which may require that professionals master a wider diversification of AD tasks.

As a result, the ADLAB PRO course materials include information about AD software and provide learners with the possibility of experimenting with an AD editor (ADLAB PRO, 2017c, 2017d). As explained by Mazur and Chmiel (2021: 63),

on the basis of the little prominence given to technology in existing AD courses, the consortium decided to feature a free online AD editor on its website to make it easier for trainers to incorporate the technological component into their classrooms.

As previously stated, AD software does not appear to be a fundamental tool for practitioners. Several professional audio describers state that they do not use specific AD software. They either write their audio description scripts using Word, or they may use some subtitling software to identify the AD in and out times: they can easily generate timecodes for the AD script. American audio describer and author Joel Snyder (personal communication, 2020) explains that he never uses AD software (either in producing professional audio descriptions or in training audio describers) and that Word and Windows Media Player are sufficient to create an audio description script. In the USA, some audio describers might use a variation of captioning software which enables them to add timecodes by pressing a button and therefore to speed up the scripting process. Snyder mentions some free and some professional Audio Description Writing Software such as LiveDescribe, SwiftAdept, Anglatècnic, together with subtitling software such as Aegisub and Subtitle Workshop which may be used to create descriptions. According to Snyder, the advantages of such software tools are that users can easily identify and insert the timecodes (i.e., the exact time in the video when the audio description is supposed to start and finish) in the script. Nevertheless, he believes that the audio editing and voicing of the descriptions must be done by professional speakers and audio editors, not by machines.

Within a European context, professional audio describer, broadcaster and scholar Louise Fryer (personal communication, 2020) explains that “professionals in the UK originally used Softel’s ADEPT” but since that is “no longer supported, . . . some use Starfish, SDI Media
Audio description software tools

uses their own. It is all expensive so students generally use subtitling software such as Swift. Currently UCL is using Stella Umbrella which began as subtitling software but is being developed for AD”. As regards theatre AD in the UK, Matthew Cock (VocalEyes) explains that their describers script the AD in Word. VocalEyes “are starting to do AD for video” and “most describers are using either Reaper, or YellaUmbrella’s Stellar software” (Cock, personal communication, 2020). In Spain, Catalan describers use Figertext by Anglätécnic, while other describers use word processors (Matamala, personal communication, 2020). In Poland, audio describers use Word and a video player, sometimes with “EZ Titles or any other subtitling software” to create an audio description script, which will then be recorded by a voice talent in a studio (Jankowska, personal communication, 2020). AD trainers such as Mazur also prefer working with Word files, without any specific AD software (Mazur, personal communication, 2021). In The Netherlands, Mereijn van der Heijden from Sound-focus explains that they do not use “dedicated AD software. Our script writers do use multiple tools. Those who are experienced with subtitling software make use of it (offline and cloud based). For some projects a simple text editor is sufficient. Recording is done in the sound studio” (personal communication, 2020). As regards the Italian context, professional audio describer Vera Arma (CulturAbile Onlus, member of the Artis Project) explains that audio describers either use Word or professional subtitling software such as WinCaps. These are useful to synchronise the audio description to the video (Arma, personal communication, 2020). Other professionals, both for screen AD and theatre and opera AD, may use the dedicated AD software AD Maker (see Arma, 2014; Lugli, 2014) or subtitling software. Elena Di Giovanni also mentions Oona Create – a tool which was not designed specifically for audio description, but which she finds useful for the scripting of screen, opera and theatre AD (Di Giovanni, personal communication, 2021). Other academic and vocational courses only use a video player and Word or free subtitling software such as Aegisub or Subtitle Workshop. On the other hand, some professionals work with AD software tools specifically developed by their clients (Catalano, personal communication, 2020; Novelli, personal communication, 2020).

A video on “Software” included in the ADLAB PRO materials (Module 2, Unit 3) contains interviews with various academics and professionals from different countries (ADLAB PRO, 2017d). Susanne Jekat, from Zürich University of Applied Sciences, mentions three different commercial software products: Starfish, Swift ADePT and MAGPie. She points out the main advantages of using dedicated AD tools: “software is a good possibility to have more than one people working on audio description, to solve technical problems and to export the audio description in an adequate format” (see www.adlabpro.eu/coursematerials/module-2/unit-3/).

The previous overview has shown that, although AD software is not frequently used by either trainers or professionals and although a video player and a Word file might be sufficient tools to teach and produce audio descriptions, a range of software programmes is indeed available. The following section describes some of the programmes mentioned by professionals and trainers.

3. Audio description software tools: an overview

This section provides an overview of AD software tools: several applications are available on the market (with subscription), while there are few free online software packages. Software can be offline (a programme which is downloaded onto the user’s computer and does not require an internet connection when producing the AD script) or cloud-based (the software
requires an internet connection, may allow access by several people at the same time, which also facilitates the sharing of files and collaboration-revision). AD software packages provide the possibility of delivering the AD lines in different ways: either through synthesised speech and/or through the recording of the describer’s own voice. In her 2016 textbook on Audio Description, British audio describer and scholar Louise Fryer dedicates a section to screen AD software, illustrating how such programmes work, what their specificities are and providing exercises. Fryer (2016: 75) points out that

the software programmes may vary in specifics, but tend to work along similar lines. The scripting software has a dedicated window allowing you to view the source programme at the same time as the window in which you write your script. . . . Hitting a specific function key as the video runs will automatically open a new scripting “box” capturing the timecode at that instant, which will be displayed in the scripting window. This is your “In time”, i.e., when your new AD utterance begins.

Fryer states that one of the advantages of AD software is that users do not have to set the exact timecodes manually, because the software does it. An AD software (like a subtitling software) allows its users to time-code the AD units, that is, to create a script containing the exact time in the video when the description starts and when it ends. The software automatically inserts the timecodes. This results in an easier and faster creation of the script, since “Timecode facilitates script creation and synchrony in screen AD” (Fryer, 2016: 85).

The course materials created by the ADLAB PRO project are also a very useful source of information on AD software for trainers and students. Unit 3 of Module 2 of the Course Materials (www.adlabpro.eu/coursematerials/module-2/unit-3/) contains videos, slides, transcriptions of the videos and PDF documents providing information on how to use AD software. The course materials focus on the software Fingertext, but also mention other tools. For instance, a video in the additional material – “Software” – contains interviews with some professionals and scholars on their use of software (see ADLAB PRO Course Materials – Module 2 – Unit 3 – Additional videos: www.adlabpro.eu/coursematerials/download/module_2_unit_3_addi
tional.zip).

The following sections will briefly describe some of the AD software tools mentioned by trainers and professionals. Basic information and some references will be provided for each tool. It is worth pointing out that AD software constantly needs updates and that new versions are often created, due to the rapid developments of technology. This is one of the reasons why AD software is seldom freely available or easily found on the web. However, the creators of such tools are often ready to provide information, to solve technical issues when these arise and to improve their software thanks to their users’ feedback.

3.1 Fingertext: an AD editor by Anglatècnic

The audio description editor Fingertext is a professional software used by Catalan audio describers (Matamala, personal communication, 2020). This tool is Spanish but has software and online help in English. With Fingertext, audio descriptions can be created both by recording the describer’s voice and by using TTS synthesis (a synthethic voice). The describer can obtain and insert in the script the timecodes (TC), that is, the “Time in” and “Time Out” of the description text, by clicking on a button. The website of Anglatecnic (www.anglatecnic.com/
project/fingertext/#editors[3] contains a detailed description of the main features of the AD editor *Fingertext*. Some of these are:

- fast time code (TC) edition;
- graphical representation of the programme’s audio (waveform). It enables the visual location of the place where the audio description segment can be inserted;
- edition of texts, comments and TCs for each audio description segment. This allows one person to prepare the script (text, comments and TCs) while another does the recording;
- segment audio through voice synthesis or recorded from the microphone connected to the computer;
- automatic simulation of the audio description (whole programme or just the selected audio description segment with the previous and following offset);
- it allows the generation of a full audio track of the audio description in MP3 or EBU WAV (BWA V) format, with or without mixing with the main programme audio (when the audio description is mixed the main programme audio is dipped accordingly);
- it can export NAR and ESF audio description files, BWA V audio file and TXT script file.

*Fingertext* is described and included in the course materials of the ADLAB PRO Project (Module 2, Unit 3). The video “Fingertext: AD editor by Anglatècnic” shows how to create an audio description script using this software tool (ADLAB PRO Course Materials – Module 2 – Unit 3 – Additional videos: www.adlabpro.eu/coursematerials/download/module_2_unit_3_additional.zip).

Moreover, Module 6, devoted to technology, contains a section called “Practice your skills” and a link to the editor *Fingertext*. Users will be asked to request a demo version of *Fingertext*. They will receive a short demo and a video, with temporary username and password and they can then use the software (https://test.anglatecnic.net/acm/sm/demo.php). Some video clips are available for practice. The online editor is called “ImAc Web Audio Description Editor” and requires a fast internet connection. The web editor also contains a user manual (Web AD Editor User Manual). Students who want to practice using this editor should read the user manual available online and read and watch the ADLAB PRO video “Fingertext: AD editor by Anglatècnic”.

### 3.2 Stellar by yellaUmbrella

The software *Stellar* is an AD editor developed by yellaUmbrella. Information on *Stellar* is available on the website: www.yellaumbrella.tv/stellar/. The website explains that Stellar is “A browser based Web App for working with timed text and media” and that it is “a Widget based application”. In order to have access to Stellar, users need to download and install a programme called Nebula on their computer and set up a Nebula account. *Stellar* only works with the browser Chrome and a fast internet connection. The software has many tools or applications. It allows its users to create and edit subtitles, scripts or transcriptions and audio descriptions. Its users can select and purchase the tools (i.e., applications) they need and decide whether they want a license for one person only or for several people who can access the files at the same time, sharing videos and texts. The AD tool allows to create audio descriptions with synthetic voice using “any available cloud voice solutions”, that is, “over 400 voices in 25 languages”. It also contains a voice recorder which enables the audio describer to record their voice, and displays the AD script in a text box. The description can be modified and rewritten and the synchronisation with the video can be adjusted. Information about *Stellar*,
with instructions and videos, can be found here: https://nebula.yella.tv/help-stellar-homepage/. A useful video provides “An overview of the Stellar time text editing system”. The software Stellar has been used by trainers at the Universities of UCL (UK) and Malta. An MA student at the University of Malta created the audio description of a recorded theatre performance liaising with the creator of this software (Cremona, 2020; Spiteri Miggiani, personal communication, 2020). As explained by Cremona (2020: 47),

The layout of the application allows the audio describer to view four sections simultaneously: the media player, where the video being audio described is viewed; the voice recorder, where the script as well as the wave form and volume of the audio describer’s voice is displayed when recording; the text box, where AD segments are created and the timeline that shows the wave form of the video’s audio and indicates the position of the AD segments alongside the shot changes. The layout of these sections can be modified in terms of position and size according to the user’s preferences.

### 3.3 Free online AD software: YouDescribe

YouDescribe (https://youdescribe.org/) is a project developed by The Smith-Kettlewell Eye Research Institute (San Francisco, California). Its creator is Dr Joshua A. Miele, Associate Director of Research and Development, The Smith-Kettlewell Rehabilitation Engineering Research Center on Low Vision and Blindness. The aim of YouDescribe is to enable people to create audio descriptions of any video available on YouTube. Users can add audio description to YouTube videos, online, recording their own voice. The website (https://youdescribe.org/) contains several videos with audio descriptions created by users.

Anyone can watch the audio described videos, while only online registered users can create and post audio described clips. As explained in the General Information about YouDescribe, “Anyone can play or describe YouTube videos for free using YouDescribe” (https://youdescribe.org/support/about). You can play a YouDescribe video without being a registered user, just by going to the main search page of YouDescribe and selecting from a list of clips. On the other hand, if you want to create audio descriptions, you have to register by logging in to an account using a Google ID. The website contains a help and support page, explanations and video tutorials.

A very helpful source of information about YouDescribe are five video tutorials contained in a YouDescribe playlist available at www.youtube.com/playlist?list=PLNjrbl_nyy9sjqZ-Wcn6sX868i9KtdNrt. When users record their own descriptions they can choose between two options on how “their voice will appear in the video”. They can decide which method to adopt: “inline description” or “extended description”. Miele, creator of YouDescribe, explains how these differ:

Inline description is where the video keeps playing and you have to squeeze your description into the available spaces in the video so that you don’t interrupt what’s going on. . . . Extended description is where you actually have the video pause for the duration of your description and then the video resumes. That’s really useful for when you’re trying to add some description where there’s absolutely no space in the soundtrack, but you need to say something in order to make the video accessible.¹

---

¹ For more information, see the video tutorials available on the YouDescribe website.
Miele points out that it is important “to preserve the flow of a video” and that the second method (extended description, pausing the video) should be used only in extreme cases. The tutorials and explanations provided by YouDescribe favour the “traditional”, “inline description” method and some of the videos uploaded on the website are described by this technique, with no interruption of the video. Nevertheless, several of the audio described YouTube videos available on YouDescribe actually adopt the method in which the description is inserted by pausing the video (“extended description”). Fryer calls this technique “‘pause for description’ method” and points out that “This means that the description is not integrated with the soundtrack and the describer is always required to post-describe” (Fryer, 2016: 80). This method is seldom used in real, professionally-produced audio descriptions, since it would disrupt the flow of the audiovisual product.

For further examples of AD software tools see the Appendix.

4. An AD experiment: the project CinemAccessibile and the software Tell Me What

CinemAccessibile was a collaborative project which was carried out in 2015 in Torino. It involved various participants (the University of Torino, the disability delegate, the Dept. of Foreign Languages and Literatures and Modern Cultures, the MA in audiovisual translation MAVTO, Museo del Cinema di Torino, IRIFOR Macerata, Neon Video) and supported by the Unione Italiana Ciechi e Ipovedenti Onlus – Torino (Italian Organization for Blind and Partially Sighted People Onlus – Torino) and Istituto dei Sordi di Torino (Institute for Deaf People – Torino). Three films about disabilities were audio described and subtitled for the deaf and the hard of hearing and were screened in a cinema during the CinemAccessibile festival (1st-2nd December 2015). Two of these films – Il lato positivo (Silver Linings Playbook) (D. O. Russell, USA, 2012) and Marianna Ucrìa (R. Faenza, Italy-France, 1997) – were audio described for the blind and the visually impaired by students and ex-students of the MA in AVT (MAVTO). The film Dancer in the Dark (Lars von Trier, Denmark, Germany, Netherlands, 2000) was audio described by IRIFOR Macerata.

This section illustrates how the online AD platform Tell Me What was used by the author with a group of MA students to create the audio description of Il lato positivo (Silver Linings Playbook) and Marianna Ucrìa. TellMeWhat was developed by IRIFOR Macerata (by Mirko Montecchiani and software developer and accessibility advisor Christoph Damm). The creators of the software were always available to solve technical problems during the audio description process. Although Tell Me What is no longer available, an account of this experience with MA students might provide insights into how AD technology can be useful to create screen audio descriptions, for collaborative work and in AD training. Moreover, it can highlight the possible applications, pros and cons of AD software and Text-to-Speech (TTS) systems.

CinemAccessibile can be considered an “experiment” in audio description for various reasons. Firstly, because we employed a newly-developed AD software (Tell Me What) for the first time. Secondly, because the audio describers of two full films were not professionals, but students and former students supervised by their lecturer and a blind collaborator/co-author. Thirdly, because the audio description was created with the active participation of a blind person who was not simply an end-user but also an active participant and co-author of the AD script. The musician Luigi Mariani (pianist and orchestra director) was involved during the AD training in class and in the process of writing and revising the AD script. A further experimental feature of this project consists of the fact that, while professionally produced audio
descriptions usually involve one or two describers (with an optional revision by a blind collaborator), in our case, several students, a blind collaborator and the lecturer worked together as a team. Both audio descriptions were supervised and revised by L. Mariani and V. Minutella. CinemAccessibile is thus an example of active cooperation with blind and visually impaired persons in AD teaching and AD scripting and could be considered an example of “participatory accessibility” (Di Giovanni, 2018). Fourthly, this project can be regarded as an experiment because the final audio description was voiced artificially (Text to Speech synthesis), whereas the audio description of films is usually delivered by a human voice (or a professional voice talent).

The online platform Tell Me What enabled the audio describers to carry out a series of actions. For example, they could: watch the video (through a video player placed on the upper-left side of the screen); play, pause, rewind and fast-forward the video; type-in the descriptions in a text box placed next to the video (on the right-hand side of the screen), thus creating the AD script; immediately listen to the descriptions, spoken by a synthetic voice. The platform also allowed its users to carry out the following tasks: inserting the exact time by clicking on a button which would pause the video and automatically capture the timecode; synchronizing the description with dialogue and important elements in the soundtrack; avoiding overlap with dialogues and meaningful noises or music by choosing the right moments in the video and moving the description forward or backwards by modifying the timecode manually, if necessary. One of the advantages of using the platform Tell Me What was also the possibility of listening to the AD merged with the soundtrack of the film. The describer was able to write the description, then play the videoclip with the AD (spoken by an artificial voice) and check whether the description was too long or too short. By playing the video of the film with the added audio description already merged with the soundtrack, it was possible to immediately identify inaccuracies, overlaps, lengthy or redundant information, repetitions, unclear descriptions, or other problems, and promptly rewrite the description, creating a new AD script. If the description was too long, adjustments could be made by changing the text, rewriting or shortening it, by choosing a faster delivery speed or making the description start a little earlier, if necessary.

The audio description process of Il lato positivo will be described in the following text. Each student had a personal account. The tutor also had access to a shared account (the Master’s account) which was used during the class sessions and where the final version of the film was uploaded. This system allowed more than one user to work on the same file/project. After some introductory lessons on audio description and on how Tell Me What worked, each MA student worked individually on the first ten minutes of the film. Each student had a personal account and could work on their own video file, create his/her descriptions and then export the audio or video file and/or a text file (the script, with timecodes automatically inserted by the software). The audio files of the students’ descriptions (the soundtrack of the film mixed with the audio description) were then sent to the trainer and the blind collaborator, who could thus listen to them and provide feedback. We then planned some sessions in class, to which the blind collaborator/co-author took part. We watched the scenes several times, with the various audio descriptions and we commented on them. We noticed the differences between the descriptions and compared our viewing experience and our “perception” of the key elements. Our blind collaborator provided comments about style, clarity, vividness and density of information, speed, and so on. The blind collaborator’s feedback on the students’ descriptions of the first ten minutes of the film and on key moments/scenes allowed us to understand his needs and helped us in the selection of the relevant information which had to be added and especially of the irrelevant information which could be trimmed from the script. The sessions in class
with the MA students and the blind collaborator were meant to share our viewing experience, our interpretation of the images and to produce a final script which would convey the essence of each scene. A constant comparison between different versions and the feedback, questions and comments provided by our blind co-author helped the students to improve and rewrite the description and to understand the approach they needed to use in describing further sequences. By the end of the first class session, a “final” version of the AD script for the first ten minutes of the film was agreed upon and written in the software programme, in the Master’s account.

After that session, each student worked individually on a different part of the film, producing their AD script and audio track delivered by a synthetic voice. The project manager, Christoph Damm, merged all the files, so that the whole film was available, audio described, on the MA account. The audio description of the full film was revised and edited by Mariani and Minutella using the software Tell Me What. Further class sessions were organised in order to discuss key elements and problematic scenes or issues. A long rewriting process ensued, guided by the blind collaborator, through group work. The final revision and editing was carried out by the lecturer and the blind co-author. Further amendments were made using the software. Key scenes were watched several times, and changes inserted through Tell Me What, until the final, definitive version was agreed upon. The aim of the final revision was to make sure that the audio description was cohesive and coherent, that there were no incongruities or unclear parts, that relevant information was selected. The revision was also aimed at reducing the length of the descriptions by trimming some words, as well as expanding some parts by adding details which it was assumed a blind person would need in order to understand a situation, a sound or a dialogue line.

A similar approach was used in the audio description process of Marianna Ucrìa, whose authors were two former students of the Master programme with previous training and experience in audio description. No class sessions were organised, but online meetings and feedback were constantly provided. The describers worked individually after having agreed on a shared but flexible method (to ensure cohesion and coherence). Each audio describer checked the colleague’s work, the first ten minutes of the film were listened to and revised by Mariani and Minutella, then the describers proceeded with the description of the whole film. The blind collaborator/co-author and the trainer then watched the film with the audio description delivered by the synthetic voice, through the platform Tell Me What and could thus immediately make adjustments by using the software. The aim of the revision was to reduce the text, avoiding overly long, detailed, “heavy” descriptions that would tire blind users, and to let the blind audience hear sounds, music, silences that in this film “describe the space” (Mariani, 2016: 411).

Notwithstanding the several positive aspects of this technical tool, it also had some negative elements. Since Tell Me What was an online platform, its functioning depended on a fast internet connection and it was not possible to use it offline. The main shortcomings of working with this software were technical problems such as poor internet connection on some occasions and several pronunciation/intonation issues (the synthetic voice would not pronounce some words properly). Nevertheless, regardless of the technical difficulties and the far-from-perfect artificial voice, the final outcome of the experiment was appreciated by the blind audience (a survey was carried out by IRIFOR Macerata after the screening of the films). Moreover, some reception studies have shown that, although blind and partially sighted people prefer human voices, they are accustomed to synthetic voices and accept their use, especially in more recent years as technology has made progress and artificial voices have a better quality. For instance, commenting on a survey on Text-to-Speech synthesis and natural voices carried out among blind and visually impaired people, Szarkowska points out that although “TTS AD” was not “the preferred solution for the reading out of the AD script”, many respondents found
it “acceptable either as an interim (95%) or an alternative (58%) solution to traditional AD” (2011: 156). Similarly, a survey carried out in 2015 by Fernandez Tomé and Matamala on artificial and human voices revealed that “94% of participants considered text-to-speech AD to be an alternative acceptable solution to human-voiced AD” (2015: 180). The scholars conclude that end-users accept artificial voices, although they prefer natural ones. (Fernandez-Tomé & Matamala, 2015: 181; see also Matamala, 2016; Szarkowska, 2011; Walczak & Szarkowska, 2012; Szarkowska & Jankowska, 2012).

Although Tell Me What is not available anymore, it proved useful for creating screen AD for the project CinemAccessibile and also in AD training at master’s level. It was used during short modules on AD at the Universities of Turin (Italy), Universitat Jaume I (Spain) and Universidad de Las Palmas de Gran Canaria (Spain). Some MA dissertations also focused on how this software could help to create audio descriptions of film logos (Novelli, 2018) and of the Netflix Series Orange is the New Black (Sciacero, 2018). On the CinemAccessibile experiment see also Minutella (2016), Montecchiani and Damm (2016), Romano (2016), Pavone and Simonigh (2016) and Biscaro and Tormena (2016).

5. The software AD Author: a new development

The new software tool AD Author has been recently developed by Christoph Damm (IT Consultant and Accessibility Manager), drawing from his previous experience with the online platform Tell Me What and from collaboration with AD professionals and blind and visually impaired people. AD Author is primarily designed to add audio descriptions to video content and is part of an AD Suite (www.audiodescription.info/). According to Damm, with the help of this software “teachers and content providers, movie makers, YouTubers, distributors or platforms can create their own audio description without the need to involve technicians, voice over artists and recording studios” (personal communication, 2021). This would make the creation of audio described video content easier and more widespread. Damm believes that AD Author would be useful also
for teachers (from various fields) who want to make their video lectures/materials accessible to students with visual impairment, without the assistance of a technician. In his words, “educators can just quickly create content without investing a lot of time and money and resources making accessibility” (Damm, personal communication – EASIT interview video, 2021).

Information on *AD Author*, instructions on how to use the software and a free demo version can be accessed and downloaded from the website www.audiodescription.info/software/adauthor/. The demo version has no limitations in terms of time, but users cannot write more than five descriptions for each clip. Despite this limit, the free demo version can indeed be used to test the software and also for an initial training of audio describers, when practicing audio description with very short clips. If, on the other hand, longer projects are required, the full license with all the tool’s functions is needed.

The following part of the chapter illustrates how *AD Author* works once the software is installed on the computer. Figure 20.1 shows the application and how to create a new project defining all the settings. To create a new project (i.e., to audio describe a video clip), click on the red icon “Add new Project” (on the top left side of the screen). A window opens (Project info). Select the video source by clicking on “browse video files”; select the video file from your own computer, choose a “Project name”, then select a “language” and a “default narrator” (a synthetic voice) from the ones available on your computer. Other important settings are the “project type” (you should select “script”) and the “default speed” (which is usually 0). You can then choose where to store the project (“Project location”) by selecting a folder in your computer. The project should then be saved by clicking on the “Save project” button.

The video will be saved and the editor with the “new Project” will appear (see Figure 20.2). *AD Author* contains a video player (top left), a wave form (under the video, showing when there are voices and noises/silences), the video volume and audio description volume, the buttons to play, pause, fast forward, rewind and stop the video as well as the window box where you should write the descriptions (on the right side of the screen). The software allows you to find the right position for audio descriptions by browsing and watching a video (noting when

![Figure 20.2 New project](image-url)
there are pauses between dialogue), but also by finding silent spots observing the audiovisualisation wave form. When you want to add a new description, you should pause the video and click on the red button with the symbol “+” (placed on the lower part of the screen, in the middle). Figure 20.2 shows the software with the new project opened.

While you watch the video, you can insert a description by pausing the video and pressing the “+” button. The window box “New description” will appear (see Figure 20.3). The time-code (i.e., the “start time” of the description within the video) is automatically inserted by the software. You do not need to worry about setting the “time in” and “time out” of the description, since the software does it automatically and includes them in the final script. The “out time” (timecode when the description ends) will be automatically set by the software.

The software allows users to create audio descriptions in three different ways: by using speech synthesis (“TTS” – Text to Speech); by importing previously-recorded professional audio files (“Sound Files”) or by recording their own voice (“Recordings”). These methods are explained here.

The first way to add an audio description text is to use the Text to Speech translation (TTS) (see Figure 20.4). It is a system which transforms the text written by the user into a speech uttered by a synthetic voice. Users write the description in the specific “Description text” box, save it and can immediately listen to the text delivered by an artificial voice, checking whether it fits in with the dialogue or overlaps with it and whether it is cohesive and coherent with previous and following dialogue.

AD Author contains one synthetic voice for each language by default. The synthetic voices are those available to the user’s computer, though other voices can be added. The demo version of the software also includes a seven-day-trial vocaliser.

When using the TTS modality, users can prepare an audio description script in two ways. They can either write the text in the “Description text” box, with no further comments, or they can insert some extra comments. If the audio description script is going to be delivered by a voice talent, the audio describer can add some extra information which can be useful for
Audio description software tools

the professional speaker who will later voice the audio description in a studio. For instance, describers can type the words immediately preceding the audio description in the “Text before” box (see Figure 20.4). This gives the voice talent who will record the audio description a cue as to when the utterance occurs (immediately after specific words). The red buttons on the right-hand side, next to the “Text before” box and above the “Description text” box, can be used to insert symbols in the script (see Figure 20.4). They provide additional information and suggestions on how to deliver the lines (S stands for fast; S+ stands for very fast; a colon stands for loud), they signal the presence of pauses . . . , of noises (NOISE) within the description, of a change of scene within the description (*) or of other elements in the utterance. These extra symbols and information will be automatically included in the final script generated by the software, so that the voice talent is helped in delivering the lines. However, during the Text-to-Speech voicing of the AD, the speech synthesis delivers the lines at a normal pace and volume, without taking those symbols into account. Example 1 shows a script file containing the audio description text with timecodes, whereas example 2 shows a script file containing extra information for the voice talent who will deliver the lines (the “text before” is in italics, within inverted commas, whereas pauses in the delivery of the audio description are signalled with “. . . ”).

Example 1: an AD script file generated by AD Author when no extra information is added.

00:00:11.1
“. . . ”
Ascolta,
00:00:12.9
“. . . ”
Example 2: an AD script file generated by AD Author with extra information.

00:00:11.0
“... Escucha”.

Ascolta. Questo non è un racconto, sono molti.

00:00:17.6
“... Son cuentos”

Sono storie di nessuno, ma possono appartenere a tutti.

00:00:23.5
“... Hablan”

Parlano di fortuna, ... amore, ... oro, ... e morte.

An alternative way to add an audio description (Figure 20.5) is by recording your own audio files through an integrated audio recorder, using the inbuilt recording function. The user decides where to add the audio description, presses the “+” button, selects “Recordings” and starts recording his/her own voice.

Figure 20.5 Adding a description recording your own voice
By using these two methods (TTS and Recordings), audio describers can obtain immediate feedback on how their description fits in with the video file with the original dialogue and whether the description is too short or too long, whether it overlaps with the dialogue or meaningful music or sounds and thus requires some adjustments. Changes can be immediately made to the script, which can be rewritten at will.

A third method (Figure 20.6) is to add a pre-recorded audio file (if the audio description has been previously recorded). A professional recording by a voice talent is obviously the best way to provide accessibility in a way the user can relate to the most. AD Author enables you to choose those recordings, which can be applied to the video at exact time stamps. This function might be used by professionals, but perhaps it is less likely to be used for AD training, since audio description tracks are not easily available.

AD Author also enables its users to create audio subtitles. When the film contains utterances in a different language (not the language of the audio description) which are subtitled into the target language, then these subtitles need to be accessible to blind or visually impaired people, otherwise they would only hear the foreign language (original audio). This is called audio subtitling (Matamala in this volume). The audio description of the subtitles needs to be recorded over the original dialogue in the foreign language, though usually the audio description does not completely cover the original voice and starts a little later, in a kind of voice-over. AD Author allows us to record the audio description over the original dialogue with TTS, adjusting the volume of the original audio track and that of the audio description at will. Figure 20.7 shows how audio subtitles can be inserted through the software. It is worth noting that adjustments to the “time in” of the description can always be easily made by changing the “Start time” in the box on the top left side of the window. For instance, if the description starts late and we want to place it a little earlier, we can modify the seconds or the frames (as is shown here, 00:03:59:10 stands for 00 hours, 03 minutes, 59 seconds, 10 frames). Changes can be made in terms of seconds or frames.

Figure 20.6  Adding a description using sound files
AD Author can be a useful tool for trainers and professionals. A professional audio describer can send the voice talent a script (the exported word file containing the audio description text, with the time-in and time-out of each description) as well as the recording of the rehearsals (either performed by synthetic voice and/or recorded by the audio describer). The audio file (if recorded by the audio describer) could also simplify the task of the speaker, as it can provide information on how to pronounce specific names or words without the need to add the pronunciation in the script or in a separate file. When explaining how to write an AD script, Fryer notes that “As the voice talent may also be unfamiliar with the source programme it may be useful to include pronunciation advice, particularly for character names or locations” (2016: 79). By exporting an audio file of the AD script (recorded by the audio describer), there might be no need to include pronunciation advice in a word file.

By the end of the script writing process, the AD script can be exported as a .docx file with timecode information. If nothing was written in the box “text before”, then the script will look as follows, where “...” stands for dialogue preceding the audio description.

Example 3: AD script file generated by the software AD Author

00:00:15.4
“...”
Si divora la fetta di torta.
00:01:11.0
“...”
Un pugno colpisce l’uomo.
00:01:17.0
“...”
Un mattone sulla torta.
Alternatively, the audio describer can fill in the “text before” box with a few words from the dialogue preceding the audio description utterance. The function of these dialogue words is to provide a cue for the voice talent as to when he/she should start speaking the audio description line. The following is an example taken from the previous description (example 3), in which the “text before” was added. It means that the voice talent should start speaking at 00:01:11.0, after hearing the utterance “Che tipo di dolce?” and then again at 00:01:17.0, after hearing “Uh!”.

**Example 4: AD script generated by the software AD Author (with “text before”)**

00:01:11.0
“. . . Che tipo di dolce?"
Un pugno colpisce l’uomo.

00:01:17.0
“. . . Uh!”
Un mattone sulla torta.

AD Author also enables users to export an audio file of the audio description (.wav file), as well as video files with an added audio description track. It is possible to export the description in different formats: as an independent audio track including all descriptions at the correct timestamp but without the video or original audio; as an audio track where descriptions are mixed with the original audio, but without video; or as a full video compilation where all the audio tracks are included (original audio, original audio including descriptions, or audio descriptions only).

A few minor negative aspects that require solution must be pointed out. If describers decide to use the “Recordings” function and thus speak the description themselves, using the microphone, at present they can record their own voice, but they cannot write the description (no “description box” is available in this modality). This means that no script can be generated and exported (and thus sent to the voice talents/professionals that would record the audio description in a studio). By using the “Recordings” modality, users can only record their voice and then listen to the audio description, but they cannot produce an audio description script (write the text, read it and export it). This shortcoming has been signalled to the software creator and an updated version of AD Author will include a “description text” box when the AD is created through TTS. Another negative aspect is related to the quality of the synthetic voices available and some pronunciation problems that may arise in some languages. The software does not provide good quality synthetic voices. Delivery of the lines by voice synthesis may not be of good quality and may thus hinder comprehension. This pitfall can be solved by using different artificial voices (purchasing them).

Nevertheless, the software is user-friendly and easily accessible. The double option of using an artificial voice or a human voice is also a positive element since it provides flexibility.
This makes it an interesting tool which can be used in AD training and by teachers or companies who want to make their videos accessible. It can also be useful for professional audio describers, especially to speed up the script creation process, to insert timecodes and to provide information to voice talents about pronunciation and delivery issues (without the need to replace professional voice talents and video and sound technicians/editors). Any person who wants to make a video content accessible can also benefit from this tool.

6. Conclusion

Although dedicated AD software is not widely used by professionals or by AD trainers, it is an interesting and practical teaching tool which can also be useful in speeding up the process of creating an AD script. More specifically, AD editors can help trainees to synchronise the audio description with the soundtrack of the product they need to audio describe, to insert the timecodes, to rehearse the lines and listen to them, verifying whether their descriptions are too long or too short, whether they overlap with the dialogues or with meaningful sounds/music, as well as to create the final audio description script. For students with no previous knowledge of audio description, or who are not trained actors, the availability of speech synthesis which “utters” their script as soon as they write it allows them to immediately check the line length, the blending/interplay of the description with the surrounding dialogue and video and to make adjustments if required. Moreover, when learners are not trained dialogue writers, synchronisation and script writing might be easier to achieve using AD software. The automatic insertion of timecodes also speeds up the audio description writing process. The possibility of experimenting with both TTS and the users’ own recorded voice is another useful (and fun) tool offered by AD Author, Fingertext, YouDescribe and other AD programmes discussed in this chapter. Audio describers can rehearse their script and change it at will, until they are satisfied with the final result.

It is undeniable that AD training must primarily focus on developing writing skills, “linguistic competences”, “competences related to the content” (Díaz-Cintas, 2006, 2007; see also Mendoza & Matamala, 2019; Mazur & Chmiel, 2021) and on the quality of the AD script. It should illustrate the “pillars” of audio description and highlight issues such as conciseness, directness, objectivity, and so on, as well as the importance of selecting relevant information, using a clear, concise and vivid language. It should provide several examples of good practice as well as make reference to available guidelines and research (ADLAB PRO; Maszerowska et al., 2014; Snyder, 2014; Remael et al., 2015; Fryer, 2016; Chmiel et al., 2019; Perego, 2012, 2016, 2017; Taylor, 2019 among others). Nevertheless “technological or applied competences” (Díaz-Cintas, 2006, 2007) could be boosted by introducing the use of AD editors. What AD software can offer to both trainers and trainees is flexibility of performance, immediate synchronisation and rehearsal and the possibility of easily creating and exporting scripts according to specific templates. User-friendly AD software can be a useful tool in some AD courses during the practical part of the modules, in order to develop and test the learners’ skills. Using AD software in class, with the teacher, as part of group work and individual work, would also develop “learning through acquisition, inquiry, discussion, practice and collaboration” (Mazur & Chmiel, 2021: 52; Laurillard, 2012).

7. Acknowledgments

I would like to thank Chris Damm for his help with AD Author, for immediately solving all technical problems and for adding new features to the software. I would also like to thank
Matteo Tortone for allowing me to use images from his film *Mother Lode* (2020); Anglatècnic and YellaUmbrella for allowing me access to their tools; scholars and professionals (Vera Arma, Alice Catalano, Matthew Cock, Irene De Higes Andino, Elena Di Giovanni, Jorge Díaz Cintas, Louise Fryer, Anna Jankowska, Anna Matamala, Iwona Mazur, Nuria Mendoza, Alessandra Novelli, Joel Snyder, Giselle Spiteri Miggiani, Mereijn van der Heijden) for answering my questions and the editors, Chris Taylor and Elisa Perego, for their advice and support.

8. Appendix: Further examples of AD software

8.1 Starfish AD tool

*Starfish AD Tool, Advantage Authoring*, developed by Starfish Technologies Limited (UK) ([www.starfish.tv/audio-description-video-description/](http://www.starfish.tv/audio-description-video-description/)), is part of a range of audio description tools called *Advantage Audio Description* product range. This software enables its users to create audio descriptions by writing the scripts and recording the audio description. The website explains that “Script text can be entered manually with a standard keyboard, or imported”. Moreover, this AD tool called *Advantage Authoring* “is available as a standalone single-user workstation, or as part of a multi-user workgroup with shared database”. It can thus be used by a person only (a teacher, student or professional), or by a group of people who work in team and need to share materials and feedback (for professional or educational purposes). It is worth noting that this AD tool is a cloud-based service. This means that it requires a fast and reliable internet connection, while no software license needs to be installed on the computer. The cost of this tool depends on the frequency of use and number of users.

8.2 Video to Voice’s Audio Description Editor

*Video to Voice*’s audio description software ([www.videotovoice.com/](http://www.videotovoice.com/)) uses Text to Speech synthesis. It allows its users to work directly in their browser. Users do not need to download the software, but they need a good internet connection. They can upload their videos to the editor, watch the video using the in-built video player and type in their description. The time in and time out of each description is automatically generated by the software. The audio description script is read out by a synthetic voice. The software has three types of subscriptions (Basic, Professionals, Teams), but users can sign up for a seven-day-free trial which allows them to use the editor writing the audio description, listening to the text-to-speech delivery of the description and exporting the AD script as. srt,. vvt or. docx files. Extra functions available for professionals or teams are the possibility of sharing the scripts with the client/tutor, who can review the AD and provide immediate feedback online.

8.3 AD Maker

The software *AD Maker*, developed by Marco Stefani (E.V.M. Service) ([www.evmservice.it/Prodotti.html](http://www.evmservice.it/Prodotti.html)), is used in Italy by some professionals for creating both film and theatre audio descriptions (Arma, 2014; Lugli, 2014; Stefani, personal communication, 2021) and to deliver audio descriptions in theatres (Di Giovanni, personal communication, 2021). *AD Maker* allows the audio describer to generate a script with the timecodes of the AD strings and to record their own voice. It counts the words per minute, signals the maximum number of words that can be used in a description and works with. srt,. txt and. doc files, which can be exported. It is also possible to export Excel files.
Note

1 “Inline Description or Extended Description?”, online at: www.youtube.com/watch?v=3gPuzzZzGfo&list=PLNjrbl_nyy9sjqZ-Wcn6sX86f9KtdNrT&index=3

9. References


10. Websites

Inline Description or Extended Description? Retrieved 9th April 2021 from www.youtube.com/watch?v=3gPuzzZzGf0&list=PLNJrbI_nyy9sjqZ-Wcn6sX868i9KtdNrT&index=3
11. Filmography

