Multimedia Accessibility

Carlos Duarte and Manuel J. Fonseca

Abstract Multimedia content is growing at an increasing pace. Making this content accessible to people with impairments is not only paramount but also a growing challenge in itself. Access services for people with visual or hearing impairments have been studied and refined over the last years, resulting in standards and laws to ensure that minimum amounts of accessible content are produced. Professional content producers have the knowledge and skills required to do so. However, individuals creating multimedia content to publish on the Internet usually lack both. In this chapter, we review existing access services, comprised of subtitles, sign language and audio descriptions. We complement this with a summary of research efforts that could assist both in the production and consumption of access services. Finally, we discuss how emerging technologies and techniques, like machine learning or crowdsourcing, can help us tackle the sheer amount of access services that need to be created to ensure all have equal access to produced content.

1 Introduction

Multimedia content has become available to large numbers of the world population since the mass commercialisation of television (TV) begun. To ensure this content is also accessible to people with disabilities a variety of access services have been made available over the years. These include subtitling and sign language interpretation for people with hearing impairments and audio description for people with visual impairments. The provision of these access services is the responsibility of

Carlos Duarte
LASIGE, Departamento de Informática, Faculdade de Ciências, Universidade de Lisboa, e-mail: caduarte@fc.ul.pt

Manuel J. Fonseca
LASIGE, Departamento de Informática, Faculdade de Ciências, Universidade de Lisboa e-mail: mjfonseca@fc.ul.pt

the TV broadcasters or content providers and regulated by the laws of individual countries.

Multimedia availability increased even further with the emergence of the Web. Not only some of the content broadcasted on TV found a new medium for dissemination, but now every individual could become a producer of audio or visual content that she or he could make publicly available on the Web. This evolution raised a new set of challenges for producing and disseminating accessible multimedia. Some of the existing access services for TV could not be readily reutilised for the Web. Many multimedia players are not accessible. Individual producers are not aware of how to produce accessible content.

Furthermore, content producers, professional or otherwise, are not able to know in advance who will consume their content and under what circumstances. Thus, it is a good practice always to make content accessible to everyone. They should not consider only people with impairments (e.g. visual impairments, hearing impairments), but also other people consuming the content in conditions or contexts that limit their options[1]. For instance, users with a low bandwidth connection, in a noisy environment (e.g. a crowded pub or bar), or in a place where they can not disturb others (e.g. a library). People that are not able to hear a video will scroll past it if it is not captioned, irrespectively of the reason they cannot hear it.

In this chapter, we start by covering existing access services and some of the regulations they are subjected to. We follow with a discussion of existing multimedia accessibility problems on the Web and present existing research that aims to tackle those. Before concluding we present an overarching discussion of this topic, some future avenues and our opinion of the field.

2 Access Services

Access to multimedia, or to any other type of content, should consider the ability to perceive the content, to understand the content and to operate the mechanisms that allow the content to be reproduced. Multimedia access services focus on offering the possibility to perceive the content, and target two main user groups. People with hearing impairments that can watch the content but not hear it, benefit from subtitling and sign language interpretation. People with visual impairments that can hear the content but not see it, benefit from audio descriptions.

Another user group that can also benefit from access services are people with cognitive impairments. Although access services target content perception, perceiving the same content from multiple channels can improve the ability of users with some cognitive impairments to understand the content (Sloan et al. 2006, Khan 2010). The impact of multimedia on people with dyslexia is also under investigation although conclusive results have not been achieved yet (Wang et al. 2018, Knoop-van Campen et al. 2018).

[1] Discussed in the "Situationally-Induced Impairments and Disabilities" chapter in the first part of this book.

Finally, in order for users to benefit from existing access services, the media player itself must be accessible. In what concerns media player accessibility, a further group of users must also be considered: those with physical impairments. For access services that are provided for broadcast content this usually means that the TV set, including the remote control, must be accessible. For access services provided for Web-based content, the media player must take into account the needs of different groups of users so that the interface is perceivable, understandable and operable. In section 3 we will further discuss the issues of media players' accessibility. In the following paragraphs we elaborate on the characteristics of access services.

2.1 Subtitles

Subtitles are text provided and synchronised with multimedia to provide the speech in the media content[2]. Another commonly used term for subtitles is captions. However, in some countries, captions refer to more than the dialogue transcription, including also sound effects and possibly the speaker's identification. Depending on the platform and service, subtitles can be visible continuously (open subtitles) or the user can select to display them in the picture as desired (closed subtitles). When considering the TV platform, one common use for subtitles in countries that broadcast foreign language programmes is to subtitle those programmes when they are not dubbed in the national language. In these situations, the subtitles make the media content accessible not only for those with hearing impairments but for the general population.

Even though subtitles are the access service with higher availability, they are still far from being universal. Figure 1, published in the European Commission's "Study on Assessing and Promoting e-Accessibility" (Kubitschke et al. 2013), shows that about half of the broadcasted programmes in the EU around 2013 had subtitles (with public broadcasters having a higher availability of subtitles than commercial ones). Although the study focuses on the EU countries, it also includes data from four non-EU countries for comparison: Australia, Canada, Norway and the USA. In all of these countries, at least 85% of the programmes have subtitles, which is a number aligned to the rate in some of the EU countries (like France, Holland, Slovenia and the UK), but much better than the average of the EU. Unfortunately, we are not aware of a more recent analysis as encompassing as this one is for TV-based media or a similar one for Web-based media.

Producing subtitles for media content equates to creating a file with time codes indicating the start and stop times for presenting the subtitle together with the subtitle's text. There are several tools available to help in subtitling efforts[3]. More recently, efforts have been made to create auto-captioning services. We will discuss some of these in section 3.

[2] As defined in the Web Content Accessibility Guidelines 2.0

[3] For example, the National Association for the Deaf publishes a list of captioning tools

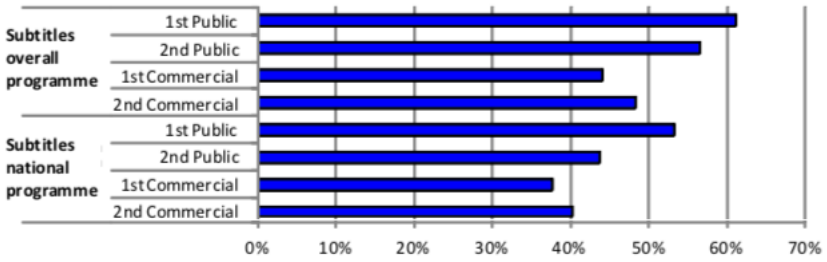

Fig. 1 Average share of programmes with subtitles in the overall programme across all countries included in the study (adapted from Kubitschke et al. (2013))

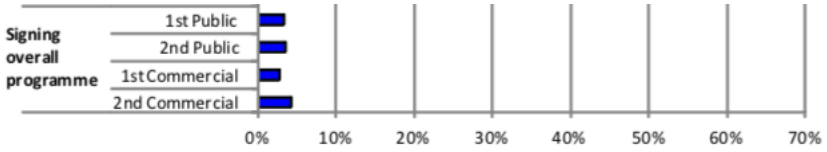

Fig. 2 Average share of programmes with sign language in the overall programme across all countries included in the study (adapted from Kubitschke et al. (2013))

2.2 Sign Language

Sign language interpretation consists in the translation of spoken words into a language that conveys meaning through combinations of handshapes, facial expressions and movements of the hands, arms and body[4]. Unlike subtitles, signing is traditionally provided in an open implementation, i.e., it is overlaid on the media content and users are not able to decide whether to enable its presentation or not. However, recent research efforts are addressing this issue and enabling closed sign language, as we will show later in the chapter.

Sign language provision is the least available of all access services. Figure 2, adapted from the study on e-Accessibility mentioned above, shows that only around 5% of all programmes have sign language interpretation, without significant variations between public and commercial broadcasters. This low percentage of programmes with sign language is common across EU (Portugal being an exception with an average of 12%) and non-EU countries.

The difficulties and costs associated with producing sign language access services are certainly not foreign to its low availability. Sign language interpretation requires capturing video of a human sign language interpreter and then overlay that video over the media content that is to be published or broadcast. Dissimilar to subtitles, which can be created without any significant infrastructure on a voluntary basis, the production of sign language interpretation is a professional task. Due to the required expressive richness in facial expressions and the movements of the

[4] As defined in the Web Content Accessibility Guidelines 2.0

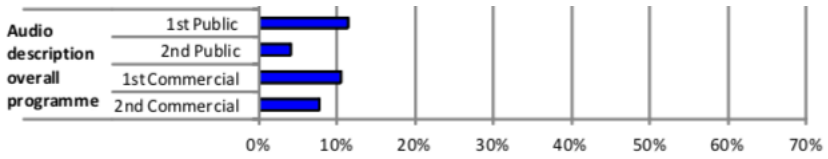

Fig. 3 Average share of programmes with audio description in the overall programme across all countries included in the study (adapted from Kubitschke et al. (2013))

hand, arms and body (Smith & Nolan 2016), solutions based on computer graphics, making use of avatars, have been unsuccessful so far (Wolfe et al. 2016), albeit having the promise of reducing the production costs for signing (Kacorri et al. 2015). Section 3 provides more details on this topic.

2.3 Audio Description

Audio description is a narration added to the soundtrack of media content to describe important visual details that cannot be understood from the main soundtrack alone[5]. This narration takes place between pauses in the dialogue and usually describes characters, their actions, scene changes or on-screen text.

The availability of audio description in EU based countries is barely better than that of sign language. Figure 3, adapted from the same study, shows that only between 4% and 11% of the broadcast programmes have an audio description available. Remarkably, the North American countries represented in this study have higher levels of audio description available, with Canada channels varying between 17% and 29% and the USA's first public channel having 85%.

The production process of audio descriptions is similar to that of subtitles. The outcome is also a file with time codes and text which means the same production tools can be used. However, while subtitles are visually displayed together with the video, audio descriptions are read aloud by a screen reader or the media player (as long as it supports audio description).

2.4 Standards and regulation

Media content is available across multiple platforms (TV, Web) and devices (TVs, desktops and laptops, tablets and smartphones). Often the content produced for one platform is also available in the other. This situation is especially true for contents created for TV.

[5] As defined in the Web Content Accessibility Guidelines 2.0

National or international laws and directives regulate TV content producers and TV broadcasters. One example is the EU's Audiovisual Media Service Directive. This directive coordinates at the EU level the multiple national legislation on audiovisual media, both for TV broadcasts and on-demand services. Existing regulations ensure that, at least, part of the broadcast content meets accessibility guidelines. However, content created by individuals to be published on the Web is not subject to regulations. Individual creators also do not have access to the same grade of tools that professional content producers have. This lack of resources makes it fundamental to provide easy ways to produce accessible media so that the ever-increasing amount of media made available is also accessible to as much as possible of the population.

Guidance on how to make accessible media is available to content creators. The Web Content Accessibility Guidelines (WCAG) provide orientation on how to make Web content accessible. They include guidelines applicable to media content, like *Guideline 1.2 Time-based Media* that specifies properties for captions, audiodescription and sign language, for example. To provide access to this content, media players also need to possess specific characteristics. A set of guidelines exists that can be applied to create accessible media players. The User Agent Accessibility Guidelines (UAAG) focus on the principles that should orient content reproducing tools, including media players. Similarly to WCAG, the UAAG include guidelines applicable to media playback, such as *Guideline 2.10 Provide control of time-based media*.

3 Multimedia Accessibility Research

Most of the multimedia content published on the Web, in particular videos, are created by individuals who are not subject to regulations and do not have access to the tools needed to make the content accessible. In this section, we present some research works that could help individuals make their content more accessible for hearing and visually impaired people.

3.1 Overcoming the Absence of Sound

3.1.1 Subtitles

Automatic Speech Recognition (ASR) software can convert speech into text, thus overcoming the absence of sound. It has evolved over the years, promising accuracies very close to 99% when correctly trained, used for dictating purposes and while using good quality microphones in a controlled acoustic environment. However, for conversations, video lessons, broadcast news, phone calls or other scenarios where we are dealing with spontaneous speech that occurs in an acoustically, lin-

guistically and structurally different environment, their accuracy worsens (Maybury 2007). However, and despite speech recognition being far from entirely satisfactory, it has been improving over the years, making it the most affordable alternative to manually creating captions for video material.

In 2009 YouTube introduced the automatic captions system, which uses Google's ASR technology[6] to perform speech transcription. This was a significant step to make videos accessible, considering YouTube is the biggest repository of videos on the Internet and the most visited by all types of users. Combined with auto-caption, Google added their translation mechanism to offer captions in different languages. In 2017 they added a new functionality to the automatic captioning system allowing it to describe sound effects[7]. More recently, in 2018, Google introduced their live automatic speech recognition (LASR) technology[8], which can automatically generate English captions for live videos, with error rates and latency approaching industry standards.

However, subtitling a video stream is not just about applying ASR. Federico & Furini (2012) presented an architecture that automatically creates captions for video lessons, by combining an off-the-shelf ASR software with a novel caption alignment mechanism to produce a timecoded transcript. Their approach overcomes the absence of timing information in the textual transcript produced by the off-the-shelf ASR by smartly introducing unique audio markups into the audio stream before giving it to the ASR software.

Shiver & Wolfe (2015) performed two studies focused on multimedia accessibility for Internet users who were deaf or became deaf at an early age. In the first study, they identified priorities for improving accessibility for deaf people. The most identified topic was the lack of accessibility to online news. Users also mentioned that they prefer captions over transcripts because the former is synced with the video. In the second study, the authors asked participants to evaluate different types of caption styles, including those generated using ASR. Results confirmed that users prefer to have videos with captions, even when they are automatically generated and consequently with some errors, than without captions.

Toledo et al. (2005) proposed a sensorial substitution system to help deaf people understand the location of sources of sound information. Their system presents visual elements in see-through glasses and captures the environment's acoustic information around the user. The sound sources are represented as Gaussian curves, where their orientation conveys the angle of the sound source, their size the power of the sound, and colour bands dividing the Gaussian depict the different frequency components of the sound. Although this solution was devised to transmit acoustic information in real time about the surrounding environment, the visual representation of the sound sources can be adapted to become another type of information that can be used to make videos more accessible.

---

[6] https://googleblog.blogspot.com/2009/11/automatic-captions-in-youtube.html

[7] https://ai.googleblog.com/2017/03/adding-sound-effect-information-to.html

[8] https://youtube-creators.googleblog.com/2018/02/updates-to-youtube-live-streaming.html

Multimedia solutions can also be used to improve the accessibility of interactions between people with different characteristics. Peng at al. recently proposed SpeechBubbles (Peng et al. 2018), a real-time speech recognition interface prototype on an augmented reality head-mounted display, to help deaf people comprehend speech from hearing users in face-to-face group conversations. Experimental results demonstrated that SpeechBubbles was suitable for group conversations between deaf and not deaf people.

3.1.2 Sign Language

Kim et al. (2004) proposed a solution for communication among deaf people in real time, which uses an intelligent avatar communication system. It supports Korean, Chinese and Japanese sign languages to overcome the linguistic barrier between different languages. Sign language translation between the languages is incorporated, as well as CG animation techniques and emotional expression methods to produce more realistic gesture images. Experimental results show that the methods could be used for sign language communications between Korean, Japanese and Chinese deaf people on the Internet.

In (de Araújo et al. 2014), the authors presented a solution for automatic generation and insertion of sign language video tracks into captioned digital multimedia content. Their approach can convert subtitles into sign language, in real-time, embedding it in the multimedia content as an extra layer of accessible content. A 3D avatar reproduces the sign language.

Even if a video or movie has a supporting video with an interpreter performing sign language, some information is still lost due to the inability to look at both the video and the interpreter simultaneously. To alleviate this issue, Kushalnagar et al. (2017) proposed a tool called closed interpreter, which can be toggled on and off by the user. Additionally, the interpreter size and its transparency and location can also be adjusted by the user. A study with deaf and hard of hearing users to find what they like about videos that come with interpreters and to identify the benefits of the offered adjustability showed that users preferred customisable interpreters over the static because they could adjust the location and transparency of the interpreter and see both the video and the interpreter.

These results are supported by a more recent study by Terrill Thompson (2018), who conducted interviews and focus groups to develop a better understanding of how persons with impairments interact with video players. The author found that for a synchronised sign language window to be effective, users should be able to control its size, position, and opacity so they can place it in the perfect position relative to the video.

3.2 Overcoming the Absence of Image

Rich and multimedia content brings severe problems to screen reader users because when the audio starts playing they have to deal with two audio streams at the same time: one delivering the original audio from the video, and the other delivering the audio from the screen reader. Moreover, there is only one physical volume control, and it is not possible to control each sound separately. A possible solution is to use two concurrent speech channels placed in a 3D space simultaneously, along the lines of Guerreiro's work (Guerreiro & Gonçalves 2016), which takes advantage of the *Cocktail Party Effect* to present several contents at the same time. For our context, one channel could be used for the main audio of the video and the other for the screen reader audio.

Flash applications (videos or interactive games) are still available on the Internet. When we think of them, accessibility is not one of the characteristics that we associate with them. In (Krüger 2008), the author demonstrated that it is possible to create accessible Flash applications for both blind and sighted users. The author adapted an existing e-learning application to make it accessible, without the need of changing its user interface. According to the author, no separated versions of the original applications were needed. Consequently, already existing Flash applications could be made accessible to blind users, without rebuilding the overall application.

Another multimedia content that poses challenges concerning accessibility are video games. In (Allman et al. 2009), the authors presented a modified version of the *Rock Band*® computer game, to allow people with no or limited vision to enjoy the game. Their solution represents visual information through haptic and audio feedback. In particular, they convey the original drumming activity of *Rock Band*® through vibrations on the upper and lower arm (for drumhead clues) and the ankle (for kick drum clue). Auditory information is used to provide feedback about the correctness of the hit. Evaluation with subjects with various levels of visual impairment revealed that they were able to master the system almost immediately. This shows that it is possible to convey visual information to users through other modalities, making multimedia content more accessible.

3.3 Multimedia Players

When choosing how to deliver multimedia content, it is important to consider options that are fully accessible. Thus, a player should at least: i) support closed captions; ii) support audio description and enable users to toggle the narration on and off; iii) have buttons and controls that can be operated without a mouse; iv) have buttons and controls labeled adequately so they can be operated using a screen reader; v) be fully functional across platforms and in all major browsers.

This list is complemented by a set of requirements presented by González et al. (2011) that should be included in a media player to make it accessible. In particular, a media player should: i) provide different alternatives together with the video,

such as captions, audio description, sign language, transcription or extended audio description; ii) provide access to all its features via mouse, keyboard, or assistive technologies (e.g. screen readers); iii) provide help and access to documentation to inform about its accessible features and how to use them.

In the last years several media players have been developed to satisfy these requirements (or at least most of them). Among those, we have Mediasite[9], Video.js[10], iTunesU[11], YouTube[12] and Able Player[13].

Able Player is a free open-source HTML5 media player, and one of the most accessible media players available. It is WCAG 2.0 Level AAA compliant and accommodates Sign Language tracks. Able Player has fully accessible player controls and, where necessary, uses ARIA[14] to expose interface elements to screen readers. It is the only media player that fully supports the HTML5 <track> element, including all five kinds of text tracks (Thompson 2015).

4 Discussion

The overall coverage of media that is made accessible is still too low as could be seen in section 2. This is true for TV broadcast, where legislation seems to be the primary driver for ensuring that broadcast media is accessible, and for the media made available online on the Web. Unfortunately, enforcing legislation on the Internet is harder. With the know-how for creating accessible media available, the initial challenge facing the accessibility community is to find ways to support the creation of accessible media in a way that requires a low amount of resources (human and material) and effort. Only with this kind of support, the **quantity of access services** will increase and approximate the needs of the population.

The following challenge facing the accessibility community is to ensure the **quality of access services**. While it is paramount that we have means to produce accessible content, the quality of this content must be enough to make it useful. The efforts mentioned above of Google's YouTube to provide auto-captions are a perfect illustration of the tension that arises between quantity and quality. While YouTube's auto-caption service is excellent from the perspective of quantity (delivering an access service for all videos in the supported languages) it still does not deliver a quality access service when the audio track in the video does not possess the needed characteristics. Consequently, it is important that the accessibility community in collaboration with each country's regulatory agencies be able to provide measures of

[9] http://delta.ncsu.edu/learning-technology/classroom-content-capture/

[10] http://www.videojs.com/

[11] http://itunes.ncsu.edu/

[12] http://www.youtube.com/

[13] https://github.com/terrill/ableplayer

[14] Accessible Rich Internet Applications (WAI-ARIA)

the quantity of access services provided (this is what has been done so far) but also of the quality of those access services.

A further challenge that impacts the accessible media production process is the source of the media. With the advance and democratisation of technology, ever more people are media content producers. The initial impact of this situation is that there are more media to make accessible by more people that do not have the expertise to do it. Further analysing this issue reveals that the amount of live media being broadcast or distributed on the Internet is also increasing. Ensuring the accessibility of live media is much more challenging than recorded media. For sign language interpretation and audio description current solutions are based on human intervention, with sign language interpreters ensuring translation of live events and trained audio describers verbally depicting events as they unfold. Automated procedures that can increase the amount of subtitles can also be used to provide subtitles for live content and thus are the more promising avenue for production of accessible live media.

In addition to the challenges faced for creating access services, a different set of challenges needs to be addressed for multimedia players, i.e., for using the created access services to ensure an accessible playback of the content. As presented in section 3.3, a variety of media players is already available capable of implementing access services. The challenges often lie in making sure that the media players themselves are accessible in addition to the content they playback. Kushalnagar et al. (2017) present one example of how that can be achieved. With their solution, a media player could use eye tracking to automatically pause a sign language interpreter when viewers are not looking at it, and resume it when they gaze again on the interpreter and speed up the replay speed of the interpreter to catch up to the current point in the video.

5 Future Directions

Machine learning based solutions are becoming pervasive across a range of domains. Multimedia accessibility is not an exception. We have presented above current research efforts and globally available services that already make use of it. The first inroads happened for subtitle generation that took advantage of evolving speech recognition systems to automate the generation workflow. These solutions are evolving to automate also the creation of audio descriptions by recognising other sounds in addition to speech (Ichiki et al. 2018). The challenges for automated sign language generation are very different. Current limitations of avatar-based signing result for the difficulty in conveying all the richness that sign language requires to transmit information. Nevertheless, some advances are underway, beginning with signing in specific domains like weather news presentation (Azuma et al. 2018, Oh et al. 2014).

Another future direction in the field of multimedia accessibility that can benefit from the increasing reach of the Web is the crowdsourcing of access services.

Crowdsourcing can also be used in combination with machine learning approaches. This is already being explored for subtitle generation (Huang et al. 2017).

Emotions that people convey while talking are valuable to produce a more engaging conversation. This is also true for dialogues that take place in multimedia content. While people usually express themselves with different tones to convey their moods, deaf individuals are not able to capture it. They rely on visuals cues which may not be enough for them to infer the emotions conveyed. Individuals with cognitive impairments also have difficulties in interpreting emotions (McCade et al. 2011). Therefore, future research on how to identify the emotions being conveyed and how to represent them in an effective way for deaf individuals or cognitively impaired individuals is an additional avenue for exploration.

6 Author's Opinion of the Field

As we already mentioned, the amount of multimedia content on the Web is increasing exponentially. Multimedia consumption is increasing, with about one third of online activity being spent watching video[15]. Correspondingly, multimedia content production is also increasing, with the latest estimates indicating that 72 hours of video are uploaded to YouTube every 60 seconds[15], for instance.

We, obviously, believe that ensuring accessibility to multimedia content is a very relevant topic. Primarily, we all have a social responsibility in ensuring that everyone has access to publicly available content, including multimedia content. Furthermore, the World Health Organization (WHO 2011) estimates that the disability prevalence on the word population is around 15% (representing approximately 1 billion individuals). This means that from a business perspective there is a huge market for accessible services, including everything that can be promoted or take advantage of multimedia delivery.

With access services becoming standardised, our opinion is that the greatest challenge for the near future is in devising ways to assist in the generation of access services for the content that individuals are now able to create and share online. Ideally, access services would be generated automatically, but that is still out of our reach. However, if we are able to provide content producers or communities of content consumers with tools that reduce the effort required for the production of access services, we should be able to see an increase in the accessibility of available multimedia content.

Another trend that must not be overlooked is the platform of choice for consumption of the content. Mobile devices are already responsible for half of the video content watched[15]. Therefore, it is essential that mobile applications that offer multimedia content are made accessible, following trends in Web media players.

[15] 37 Staggering Video Marketing Statistics for 2018 | Wordstream

7 Conclusions

The amount of available multimedia content keeps increasing at a great pace. To make this multimedia accessible to all population is a massive trial that must be addressed at all points in the production to the consumption chain. First, access services must be created to ensure that people with visual or hearing impairments can access the content. Then, playback mechanisms, which themselves must also be accessible, need to be compatible with these access services so that consumers can, in fact, access the content.

In this chapter we tried to provide an overview of the status of multimedia accessibility, describing what access services exist, their purpose and presenting a panorama of their usage, with a greater focus on broadcast multimedia. We complemented this with a summary of existing research efforts that have the potential to lessen the required resources for the production of access services or to improve the accessibility of playback tools. Finally, we discussed how machine learning, possibly supported by crowdsourcing, has the potential to improve the quantity and quality of access services.

Acknowledgements This work was supported by national funds through Fundação para a Ciência e Tecnologia, under LASIGE Strategic Project - PEst-OE/EEI/UI0408/2014.

References

Allman, T., Dhillon, R. K., Landau, M. A. & Kurniawan, S. H. (2009), Rock vibe: Rock band®computer games for people with no or limited vision, *in* 'Proceedings of the 11th International ACM SIGACCESS Conference on Computers and Accessibility', Assets '09, ACM, New York, NY, USA, pp. 51–58.

Azuma, M., Hiruma, N., Sumiyoshi, H., Uchida, T., Miyazaki, T., Umeda, S., Kato, N. & Yamanouchi, Y. (2018), Development and evaluation of system for automatically generating sign-language cg animation using meteorological information, *in* K. Miesenberger & G. Kouroupetroglou, eds, 'Computers Helping People with Special Needs', Springer International Publishing, Cham, pp. 233–238.

de Araújo, T. M. U., Ferreira, F. L., Silva, D. A., Oliveira, L. D., Falcão, E. L., Domingues, L. A., Martins, V. F., Portela, I. A., Nóbrega, Y. S., Lima, H. R., Filho, G. L. S., Tavares, T. A. & Duarte, A. N. (2014), 'An approach to generate and embed sign language video tracks into multimedia contents', *Information Sciences* **281**, 762 – 780. Multimedia Modeling.

Federico, M. & Furini, M. (2012), Enhancing learning accessibility through fully automatic captioning, *in* 'Proceedings of the International Cross-Disciplinary Conference on Web Accessibility - W4A '12', ACM Press, New York, New York, USA, p. 1.

González, M., Moreno, L., Martínez, P. & Iglesias, A. (2011), Web accessibility requirements for media players, *in* P. Campos, N. Graham, J. Jorge, N. Nunes,

P. Palanque & M. Winckler, eds, 'Human-Computer Interaction – INTERACT 2011', Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 669–674.

Guerreiro, J. a. & Gonçalves, D. (2016), 'Scanning for digital content: How blind and sighted people perceive concurrent speech', *ACM Trans. Access. Comput.* **8**(1), 2:1–2:28.

Huang, Y., Huang, Y., Xue, N. & Bigham, J. P. (2017), Leveraging complementary contributions of different workers for efficient crowdsourcing of video captions, *in* 'Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems', CHI '17, ACM, New York, NY, USA, pp. 4617–4626.

Ichiki, M., Shimizu, T., Imai, A., Takagi, T., Iwabuchi, M., Kurihara, K., Miyazaki, T., Kumano, T., Kaneko, H., Sato, S., Seiyama, N., Yamanouchi, Y. & Sumiyoshi, H. (2018), Study on automated audio descriptions overlapping live television commentary, *in* K. Miesenberger & G. Kouroupetroglou, eds, 'Computers Helping People with Special Needs', Springer International Publishing, Cham, pp. 220–224.

Kacorri, H., Huenerfauth, M., Ebling, S., Patel, K. & Willard, M. (2015), Demographic and experiential factors influencing acceptance of sign language animation by deaf users, *in* 'Proceedings of the 17th International ACM SIGACCESS Conference on Computers and Accessibility', ASSETS '15, ACM, New York, NY, USA, pp. 147–154.

Khan, T. M. (2010), 'The effects of multimedia learning on children with different special education needs', *Procedia - Social and Behavioral Sciences* **2**(2), 4341 – 4345. Innovation and Creativity in Education.
URL: *http://www.sciencedirect.com/science/article/pii/S1877042810007305*

Kim, S.-W., Li, Z.-X. & Aoki, Y. (2004), On intelligent avatar communication using korean, chinese and japanese sign-languages: an overview, *in* 'Procceedings of the 8th Control, Automation, Robotics and Vision Conference (ICARCV'04)', Vol. 1, pp. 747–752 Vol. 1.

Knoop-van Campen, C. A. N., Segers, E. & Verhoeven, L. (2018), 'The modality and redundancy effects in multimedia learning in children with dyslexia', *Dyslexia* **24**(2), 140–155.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1002/dys.1585*

Krüger, M. (2008), Accessible flash is no oxymoron: A case study in e-learning for blind and sighted users, *in* K. Miesenberger, J. Klaus, W. Zagler & A. Karshmer, eds, 'Computers Helping People with Special Needs. ICCHP 2008', Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 362–369.

Kubitschke, L., Cullen, K., Dolphin, C., Laurin, S. & Cederbom, A. (2013), Study on assessing and promoting e-accessibility, Technical report, European Commission DG Communications Networks, Content & Technology.

Kushalnagar, R., Seita, M. & Glasser, A. (2017), Closed ASL Interpreting for Online Videos, *in* 'Proceedings of the 14th Web for All Conference on The Future of Accessible Work - W4A '17', ACM Press, New York, New York, USA, pp. 1–4.

Maybury, M. (2007), Searching conversational speech, *in* 'Proceedings of the SIGIR workshop on Searching Spontaneous Conversational Speech (SSCS'07)'.

McCade, D., Savage, G. & Naismith, S. L. (2011), 'Review of emotion recognition in mild cognitive impairment', *Dementia and Geriatric Cognitive Disorders* **32**(4), 257–266.
URL: *https://www.karger.com/DOI/10.1159/000335009*

Oh, J., Jeon, S., Kim, M., Kwon, H. & Kim, I. (2014), An avatar-based weather forecast sign language system for the hearing-impaired, *in* L. Iliadis, I. Maglogiannis & H. Papadopoulos, eds, 'Artificial Intelligence Applications and Innovations', Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 519–527.

Peng, Y.-H., Hsi, M.-W., Taele, P., Lin, T.-Y., Lai, P.-E., Hsu, L., Chen, T.-c., Wu, T.-Y., Chen, Y.-A., Tang, H.-H. & Chen, M. Y. (2018), Speechbubbles: Enhancing captioning experiences for deaf and hard-of-hearing people in group conversations, *in* 'Proceedings of the Conference on Human Factors in Computing Systems (CHI'18)', ACM, New York, NY, USA, pp. 293:1–293:10.

Shiver, B. N. & Wolfe, R. J. (2015), Evaluating Alternatives for Better Deaf Accessibility to Selected Web-Based Multimedia, *in* 'Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility - ASSETS '15', ACM Press, New York, New York, USA, pp. 231–238.

Sloan, D., Stratford, J. & Gregor, P. (2006), 'Using multimedia to enhance the accessibility of the learning environment for disabled students: reflections from the skills for access project', *ALT-J* **14**(1), 39–54.
URL: *https://doi.org/10.1080/09687760500479936*

Smith, R. G. & Nolan, B. (2016), 'Emotional facial expressions in synthesised sign language avatars: a manual evaluation', *Universal Access in the Information Society* **15**(4), 567–576.

Terrill Thompson (2018), 'Media Player Accessibility: Summary of Insights from Interviews & Focus Groups', *Journal on technology and persons with disabilities* **6**, 325–335.

Thompson, T. (2015), *Video for All: Accessibility of Video Content and Universal Design of a Media Player*, second edn, Boston: Harvard Education Press, pp. 259–273.

Toledo, J., Torres, J., Alonso, S., Toledo, P. & González, E. J. (2005), SLOAS: hearing with the eyes, *in* 'Proceedings of the 4th WSEAS International Conference on Electronic, Signal Processing and Control', pp. 27:1 – 27:5.

Wang, J., Dawson, K., Saunders, K., Ritzhaupt, A. D., Antonenko, P. P., Lombardino, L., Keil, A., Agacli-Dogan, N., Luo, W., Cheng, L. & Davis, R. O. (2018), 'Investigating the effects of modality and multimedia on the learning performance of college students with dyslexia', *Journal of Special Education Technology* **33**(3), 182–193.
URL: *https://doi.org/10.1177/0162643418754530*

WHO (2011), *World Report on Disability*, World Health Organization.

Wolfe, R., Efthimiou, E., Glauert, J., Hanke, T., McDonald, J. & Schnepp, J. (2016), 'Special issue: recent advances in sign language translation and avatar technology', *Universal Access in the Information Society* **15**(4), 485–486.