Giving Voice to Images: Audio Description and Visual Impairment: Technological Solutions and Methodological Choices

Stefania Pinnelli and Andrea Fiorucci

Abstract This paper proposes a reflection on issues related to Audio Description: an additional narration track intended primarily for blind and visually impaired people of visual. Starting from the analysis about the users' needs, the paper proposes a comparison between three types of technology solutions, taking into account and evaluating, in a perspective of social inclusion and Ambient Assisted Living, their strengths and weakness.

1 Cultural Policies for Inclusion and Accessibility

According to the World Health Organization, a disabled person is any individual with limited ability or inability to perform daily activities. In ICF the Functioning and Disability are understood as "umbrella terms denoting the positive and negative aspects of functioning from a biological, individual and social perspective" [31]. The social and environment components can become the most important factors that can transform a disease in disability. Accessibility is not the creation of specific areas reserved to disabled people, but rather one mode of rethinking systems and environments, which can be used by all [31].

Stefania Pinnelli wrote the Sects. 1, 4, 5.
Andrea Fiorucci wrote the Sects. 2, 3.

S. Pinnelli (✉) · A. Fiorucci
University of Salento, Via Stampacchia,
46, 73100 Lecce, Italy
e-mail: stefania.pinnelli@unisalento.it

A. Fiorucci
e-mail: andrea.fiorucci@unisalento.it

© Springer International Publishing Switzerland 2015

B. Andò et al. (eds.), *Ambient Assisted Living*, Biosystems & Biorobotics 11,
DOI 10.1007/978-3-319-18374-9_32

It is necessary, therefore, to invest efforts in the research and development of accessible device through the perspective of a universal design. Accessibility besides to indicate the usability level of an environment, a good or a service, also refers to the virtual barriers and to the difficulties faced by different types of users in the management and use of the technologies. The cultural perspective of the accessibility concept refers, firstly, to a model of inclusive design of cyberspace, that takes account of the plurality of users' needs and gives the same services and the same interaction and participation spaces in compliance with criteria and official guidelines (www.w3c.it) and foreseeing the use of the assistive technologies. In addition to this first aspect, there is function that technologies play to improve the inclusive capacities of social contexts, turning them into participatory and interactive contexts. The design for all, also called *universal design*, began focusing on physical aspects (buildings, urban spaces, transport, health, leisure), and nowadays is extended to the digital world (computer networks and communication systems). In this perspective, accessibility is defined as "a condition for use with security and autonomy space, furniture and urban facilities, the buildings, transport services and devices, systems and media and information for people with disabilities or reduced mobility" [7]. The right to accessibility of an asset or service must be ensured equally to all citizens. It follows that the needs of people with difficulties must be taken into account in designing a strategy for the development of an egalitarian society, to avoid creating, even in "virtual contexts", new forms of social exclusions.

Within the European Disability Strategy 2010-2020 the *accessibility* is a precondition for participation in society, the second key words on which the Document points the attention is Participation: there are still many obstacles preventing people with disabilities from fully exercising their fundamental rights [1]. In the cultural sector is an important milestone was reached in 2003 with the *Resolution of the Council of the European Union on May 5* on equal opportunities for pupils and students with disabilities in education and training. It relates to the access of disabled persons to infrastructure and cultural activities and devotes specific attention to the role of new technologies in this action. More and more legal regulations in relation to media accessibility are being passed now, e.g. the Audiovisual Media Directive 2007/65/EC which requires implementation in the various EU Member States [10].

Even the world of art and cinema make use of language and means of expression that are often inaccessible or hardly accessible: for example blind and deaf people have several difficulties when they want to "see" or "hear" a film or a theatrical performance. In a society basically anchored to the visual content as a source of information, entertainment and education, people with a severe impairment or absence of vision are likely to be culturally and cognitively disabled [6]. For some time, museums and art galleries offer verbal descriptions (audio guide) of paintings, sculptures and other parties combined visual, often, a "touch tour", in order to give the blind the opportunity to access the art world, chance thwarted by the, by now too well known rule of "see, but not touch". Compared to audiovisual content (TV shows and movies) or artistic- performances (dance, theater etc.). The

social context is even less "attentive" to offer concrete answers to the needs of visually impaired people. The problems of Web data, digital data, broadcasting and cinema product accessibility, and usability for blind users have become an active research field for the past decade. In recent years there has been witnessed a proliferation of efforts aimed at making audiovisual programs accessible to people with visual impairments through audio description (AD) [19]. ICTs, domotic technologies and anything that could be Ambient Assisted Living (AAL) and Assistive Technologies (TA) are undoubtedly an excellent opportunity to improvement in the person's life, especially for those who live in situations of disadvantage, disability, illness and aging [21]. Smart houses and living environments have been designed to meet this needs becoming today an area of very active research. There are several main applications of the AAL, but one nearest to the scope of study of AD is the one to make your life easier by compensating their disabilities through home automation or other technology. AD, therefore, can be a valuable tool to meet the needs of the users with gradual or total loss of visual acuity, but also of older people that due to a deterioration of view or sometimes cognitive they need a voice/audio support.

2 AD: Characteristics and Aims

The first audio descriptive support is the human one: blind ask for significant people (family, friends, spouses) questions about visual elements. If the *whisper mode*, as defined by Snyder [23], on the one hand makes request for help spontaneous, on the other hand doesn't release the person with visual disability by dependence.

In recent years there has been a proliferation of initiatives designed to make accessible the audiovisual messages [19], through a technological process called audio description (AD): assistance and support service, composed by a set of techniques, methods and skills whose main goal is to reduce the visual deficit by making accessible any visual message with appropriate audiovisual substitute information in relation to the specific needs of potential users. It is a voiceover, aimed at describing aspects of audiovisual products that are not accessible: visual component (shares, body language, facial expressions, setting, clothes/costumes). Everything that is used to refer to the "visual world" is made accessible by a verbal description that fits between the dialogues, does not overlapping to the significant music and sound effects.

AD may include information about actions, change of scene, text appearing on screen to the descriptions of the characters, their movements and body language, the explanation of the sound effects, etc. [17, 29].

It formed by several audio comments: narrative captions that are generated in order to decode the silence moments of the audiovisual to make them accessible to the blind. Audio comments, subject to restrictions of time (time codes) required by the parent track, merge into a script (cloth) which is subsequently read by a

professional narrator and recorded by audio technician. Recordings of audio commentaries, associated with the respective silence moments of the audiovisual, give life to the new audiovisual product audio described. From a theoretical point of view, AD can be defined as a specific form of inter-semiotic, inter-modal or cross-modal translation or mediation [2, 5, 13, 16, 18]. AD, however, can be regarded as an act of translation not entirely pure, because it does not determine a direct passage between two equal semiotic codes. For this reason, scientific community [9, 27], when it relates to this particular act of translation, uses of the umbrella word *accessibility*.

OFCOM guidelines [17] drawn up by the *Independent regulator and competition authority for the UK communications industries* are the coordinates and operating recommendations to make AD. They are a result of a long process of empirical research with blind people: they are seven steps.

Step 1 Choosing Suitable Programmes for Description
Step 2 Viewing the Programme
Step 3 Preparing a Draft Script
Step 4 Reviewing the Script
Step 5 Adjusting the Programme Sound Level
Step 6 Recording the Description
Step 7 Reviewing the Recording

3 AD: Educational Potentialities

Studies about AD conducted in recent years, have focused on the evaluation of the benefits that such support may have about: **learning processes**, **mentalistic abilities** and **social inclusion** [12].

Learning processes. According to Vygotskij [30] the sensory deprivation produces a functional re-organization: blind people activate vicariant processes aimed to knowledge construction. They should follow a path much longer and harder order to build the world of objects, give them a name and give them quality and actions of which he hasn't direct experience. Language is considered the compensatory tool most functional [3] because through it blind can see what is happening. Blind people, like everyone else, build mental models to represent the world around them, abstract concepts or sequences of events and use them to give an explanation to its events to understand their experiences and cope with the new situations.

AD can: generate, through few words but carefully chosen, long-lasting and clear images in the mind of blind, in particular through the use of new vocabulary, comparisons and simple metaphors, thus facilitating learning and by improving their skills language [24]; can help to develop or acquire new visual knowledge especially for those who are blind from birth [14, 15].

Mentalistic abilities. The view, therefore, is the preferred channel to access the mental life of the other and its absence, or severe impairment, generates a delay in the development of Theory of Mind or mentalistic abilities [4, 11]. Some studies confirm that blind people have a solid verbal memory [28] with high compensators potentialities [22, 26] and that memory takes lifeblood by language because it turns dark into word.

Social inclusion. Television is one of the main sources of information, education and entertainment for citizens. Lack of accessibility can lead to the large scale social exclusion of people with vision impairments and other disabilities. Conversely AD can become an inclusive tool: it can transform a movie theater, a lounge projection, a multimedia classroom into cultural places for all. Therefore AD may promoting a sense of independence, equality and participation.

4 Implementig AD: Technological Approaches and Methodological Differences

The three approaches, that will be presented, have been chosen on the basis of three distinct experiences applied research carried out in the last 3 years, at the working Group of the Center for new technologies for disability and inclusion at the University of Salento. They are: the post-post production of the AD on the audio track of the film; the Movie Reading system and the technology of text to speech conversion.

Any work of AD, should observe three fundamental principles: (1) Respect for the work that you are describing. AD should try to preserve the atmosphere of audiovisual products while preserving the pace and adapting the description to the genre and style of the original; (2)Respect for the user. AD must consider the heterogeneity of users [14]; (3) Respect for the objectivity. The AD must not impose an emotional connotation but verbalize what appears on the screen without interpreting it.

We analyze three approaches:

Post production work. This approach involves an assembly work in the post post-production of a film already edited and on which the working group writes, calibrate, adjust and add the audio track of the AD. Construction of an audio described using that more traditional approach happens within a process of research and experimentation that provides different stages of post-editing. The process of AD is very large and complex and it requires inevitably a team work. The collaboration of a blind person is crucial. Using this approach has been realized, from research group in the 2011 the AD of an audiovisual for children: *Kirikou et la sorcière*, written and directed by Michel Ocelot in the 1998 [20].

This approach allows you to create AD very accurate and understandable by a wide audience, or for a targeted audience (for example, children as in our case), allows to treat all the elements of intonation, prosody, conceptual clarity related to

the hermeneutic dimension of language and, therefore, it is particularly suitable in AD aimed at processes of learning and promotion of language development. This approach consists in fact a single track with images, audio and AD, the product is usable on all devices, free from particular technical requirements and devoid of issues related to sync, remote controls, specific requirements of the operating system. However, it is expected a long time to manufacture, and therefore higher costs. This approach also raises a fundamental problem: the copyright. Another element that in some cases it may be considered a disadvantage, but in most cases it was considered, as our research argues, an advantage for understanding, storing information and attention [20] is the fact that the entire audience, the blind and visually impaired, uses a film with audio description.

MovieReading system. A second solution is the application MovieReading, it allows to go to the cinema and search on your smartphone or tablet subtitles (for deaf people) or tracks AD (for blind people) of the film screened in room. In fact, the process of production reduces the steps compared to the previous solution, requiring a dialogist that plays the movie and write the dialogues for AD; checking recursive comparison with the blind person to validate the quality of the message, then proceed to the registration of dialogues in a separate file for MovieReading. The recorded track is tagged by the system in a few points allowing you to synchronize the audio of the film. MovieReading "listens" to the audio of the movie through the microphone and is able to identify the exact point of the movie you are watching, synchronizing AD and/or subtitles. The AD of movies are downloadable, with a small cost (about € 2), with a specific APP for I Pad, I Phone, tablet and smartphone.

Using this system has been made the AD of "The Great Beauty", Oscar-winning film in 2014 on which we have done a workshop to test the effect on the public with visual impairment and not (http://goo.gl/BeabWA). In this experiences has been possible to assess the MovieReading App: it is a quite good solution and is an Italian product but it needs to be improved in functionality to sync the sound track audio described. The advantages of this type of solution resides in the first instance in the reduction of costs and the use of copyright the original film . In fact, the user purchases the AD and this allows him to re-use that track as many times as you want . It 'also the reduction of costs for staff involved in the development of AD, lacking the entire post post production, audio-video presentations on the film. Each AD can be done with about 1.500/2000€. One problem is that the purchased track to sync with the original track of the film, it needs a sound quality similar to that of the cinema, otherwise it loses sync, it is not always what is achievable, if we think of solutions to home television or movies in educational context. Another problem could be that the blind person can be disturbed because has one ear that hear the general sound of the film and the other the AD voice.

Technology of text to speech conversion. The third approach shows the text-to-speech audio description (TTS AD): AD is read by speech synthesis software. Modern text-to-speech applications convert text input into a speech waveform with the use of special algorithms [8], producing an effect far more natural than speech synthesizers did a few years ago [27]. Text-to-speech audio description

has several advantages. From the perspective of the audio description provider, TTS AD offers unequalled cost-effectiveness in terms of AD production in comparison with conventional methods of producing audio description. TTS AD does not require the recording of the AD script (for pre-recorded AD), nor does it incur any human labour costs for the reading out of the AD script (for live AD). Furthermore, in contrast to audio describers involved in the production of conventional AD, who need to be able to develop "the vocal instrument through work with speech and oral interpretation fundamentals" [25], audio describers for TTS AD do not need to have any particular vocal skills [27]. From the end user's point of view, TTS AD also helps many blind and partially sighted people save money, as they already have speech synthesis software at home or at work and are accustomed to using it in their daily lives. Thanks to the high quality of speech synthesis software available now in many languages, watching a film with synthetic AD can be an enjoyable and entertaining experience. For those watching the audiovisual programme another advantage is that the solution does not require access to a high-speed Internet connection because the viewer is simply offered a text file with the AD script (in .txt or .sub format) to be read out by a text-to-speech programme. The solution seems particularly attractive to those visually impaired people who live in small towns and villages and thus cannot enjoy cinema screenings or theatre performances with AD, as these are usually organized in large cities only. Furthermore, TTS AD allows spectators with visual impairments to watch films and other audiovisual programs on their own, without depending on others or being restricted to the explanations of their sighted friends or family. At the moment the most used software are Ivona, Acapela, Loquendo, Readspeaker. Ivona is the better Italian female and masculine voice. Field surveys carried out by following this model of audio description, testify that many blind people prefer a speech because it allows you to better distinguish between original audio and audio description.

5 Conclusion: The Power of Words

In conclusion we can say that the solutions of AD are varied, the choice should be made on the basis of time available, the skills available, costs and the purpose of the product that you want to accomplish of course, a traditional system of AD provides enhanced quality and responsiveness to the needs of interpretation and decoding of the person who is blind, however, this often means a reduced amount of film production and film often outdated. From my point of view if you intend to practice the way of accessibility and autonomy of the person with the highest visual deficits is appropriate to think of textual summaries of quality that would reduce to a maximum of three professionals work, dialogue writer, editor and blind person who regards the product. In fact, the common denominator among these or other solutions that you can choose to be the presence of human validator, that the blind person who verifies and checks the result. This step is almost

inevitable and what really can safeguard the quality of the final result. So thinking about an AD solution we needs to adopt a User-Centred Design approach to address the interpretation of blind people needs, involving them not only as stockholders but also as real validator.

In technology in recent years there have been several advances on the expansion of refining and humanization of synthetic speech. The use of speech synthesis, such as text-to-speech AD [27] is more advantageous for at least two aspects.

The first is related to the economic aspect: the speech with audio descriptions are much cheaper because (their) implementation process covers only the figures of the describers (descriptors which draw up the script) and the technical editing of audio. The second aspect is related to practicality: an AD read by a speech synthesizer can be more practical and easy to generate, thus allowing a more massive and rapid spread and rapid translation into multiple languages.

The audio description voice against human turns out to be more expensive and more complex, however, experimental studies [8, 27] confirm that the recording of a human voice is much favored by people who are blind, even the report of the technical expertise that requires the use of a summary with respect to the use of an audio description sound already assembled and with the human voice. The usability and ease of access to synthesis systems, therefore, is one of the most significant variables in the choice between audio description with human voice and digital synthesis. Especially for older people's audience. The most important thing is that the blind people, regardless of the means and the way in which it is made an AD, can experience all this in autonomy. AD, therefore, is emerging as a powerful instrument of culture and social inclusion, but unfortunately, still "limping" in the Italian context.

References

1. ANED Network of European Disability experts: National accessibility requirements and standards for products and services in the European single market: overview and examples, by Mark Priestley. http://www.disability-europe.net/theme/accessibility/two-reports-published-accessibility (2013)
2. Benecke, B.: Audio description: phenomena of information sequencing. In: EU-High-Level Scientific Conference Series MuTra 2007 LSP Translation Scenarios: Conference Proceedings (2007)
3. Bonfigliuoli, C., Pinelli, M.: Disabilità visiva. Teoria e pratica nell'educazione per alunni non vedenti e ipovedenti. Erickson, Trento (2010)
4. Brambring, M., Asbrock, D.: Validity of false belief tasks in blind children. J. Autism Dev. Disord. **40**(12), 1471–1484 (2010)
5. Braun, S.: Audiodescription from a discourse perspective: a socially relevant framework for research and training. Linguistica Antverpiensia, NS6, pp. 357–369 (2007)
6. Braun, S.: Audiodescription research: state of the art and beyond. Trasl. stud. New Millennium **6**, 14–30 (2008)
7. Costa et alii: Accessibility in digital television: designing remote controls. IEEE Trans. Consum. Electron, Piscataway **58**(2), 605–611 (2012)

8. Cryer, H., Home, S.: User attitudes towards synthetic speech for Talking Books. RNIB Centre for Accessible Information. Research report 7, Birmingham (2009)
9. Díaz Cintas, J.: Accessibility for all. Transl. Today **4**, 3–5 (2005)
10. European parliament: Directive 2007/65/ec of the european parliament and of the council of 11 December 2007, Official Journal of European Union
11. Fiorucci, A.: Teoria della mente e disabilità visiva. L'integrazione scolastica e sociale **12**, 224–229 (2013)
12. Fiorucci, A., Pinnelli, S.: Audio descrizione e disabilità visiva. Ital. J. Special Educ. Incl. **1**(1), 133–147 (2013)
13. Gambier, Y.: La traduction audiovisuelle: un genre en expansion. Meta **49**(1), 1–11 (2004)
14. González, L., Iorfida, E.: Cine a ciegas. http://goo.gl/Jvt6W (2002)
15. Hernández-Bartolomé, A.I., Mendiluce, Cabrera G.: Audesc: translating images into words for spanish visually impaired people. Meta **49**(2), 264–276 (2004)
16. Lavaur, J.-M., Serban, A.: La traduction audiovisuelle. Approche interdisciplinaire du sous-titrage. De Boeck, Bruxelles (2008)
17. OFCOM: Guidelines on Standards for Audio Description. http://goo.gl/9nt5v (2000)
18. Orero, P.: Audio description: professional recognition, practice and standards in Spain. Transl. Watch Q. **1**, 7–18 (2005)
19. Orero, P.: Sampling audio description in Europe In: Díaz Cintas, J. et al. (eds.) Media for All. Subtitling for the Deaf, Audio Description and Sign Language, pp. 111–125. Rodopi, Amsterdam and New York, (2007)
20. Pinnelli, S., Fiorucci, A.: Accessibilità e inclusione. Kirikù e la strega Karabà-AD: una ricerca intervento in classe. In: Parmigiani, D., Pennazio, V., Traverso, A. (eds.) Learning & Teaching with Media & Technology.ATEE—SIREM Winter Conference Proceedings. A T E E aisbl, Brussels (2013)
21. Pinnelli, S.: Ambient assisted living, innovazione tecnologica e inclusione. MediaEducation **5**(1), 1–13 (2014)
22. Roder, B., et al.: Auditory memory in congenitally blind adults: a behavioral-electrophysiological investigation. Cogn. Brain. Res. **11**(2), 289–303 (2001)
23. Snyder, J.: Audio description: the visual made verbal. Int. Congr. Ser. **1282**, 935–939 (2005)
24. Snyder, J.: Audio description-an aid to literacy. In: paper presented at Languages and the Media the 6th International Conference & Exhibition on Language Transfer in Audiovisual Media, Berlin
25. Snyder, J.: Audio description: the visual made verbal In: Díaz Cintas, J. (eds.) The Didactics of Audiovisual Translation. John Benjamins Publishing Company, Amsterdam /Philadelphia (2008)
26. Swanson, H.L., Luxenberg, D.: Short-term memory and working memory in children with blindness: support for a domain general or domain specific system? Child Neuropsychology. **15**(3), 280–294 (2009)
27. Szarkowska, A.: Text-to-speech audio descript ion: towards wider availability of AD. J. Specialised Transl. **15**, 142–162 (2011)
28. Tadic, V., Pring, L., Dale, N.: Are language and social communication intact in children with congenital visual impairment at school age? J. Child Psychol. Psychiatry **51**, 696–705 (2010)
29. Vercauteren G.: Towards a European guideline for audio description In: Díaz Cintas, J., et al. (eds) Media for All. Subtitling for the Deaf, Audio Description and Sign Language. Rodopi, Amsterdam and New York, 139–149 (2007)
30. Vygotskij, L.S.: Fondamenti di difettologia. Bulzoni, Roma (1986)
31. WHO (2001). ICF. Classificazione internazionale del funzionamento, della disabilità e della salute, Erickson, Trento