
A method to support accessible video authoring

Johana Maria Rosas Villena

SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito:

Assinatura: _____

Johana Maria Rosas Villena

A method to support accessible video authoring

Doctoral dissertation submitted to the Instituto de Ciências Matemáticas e de Computação – ICMC-USP, in partial fulfillment of the requirements for the degree of the Doctorate Program in Computer Science and Computational Mathematics. *FINAL VERSION*

Concentration Area: Computer Science and Computational Mathematics

Advisor: Prof. Dr. Rudinei Goularte

Co-advisor: Profa. Dra. Renata Pontin de Mattos Fortes

USP – São Carlos
June 2016

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi
e Seção Técnica de Informática, ICMC/USP,
com os dados fornecidos pelo(a) autor(a)

RR778m Rosas Villena, Johana Maria
m A method to support accessible video authoring /
Johana Maria Rosas Villena; orientador Rudinei
Goularte; co-orientadora Renata Pontin de Mattos
Fortes. -- São Carlos, 2016.
215 p.

Tese (Doutorado - Programa de Pós-Graduação em
Ciências de Computação e Matemática Computacional) --
Instituto de Ciências Matemáticas e de Computação,
Universidade de São Paulo, 2016.

1. alternative content authoring. 2. accessible
video production proces. 3. video accessibility. I.
Goularte, Rudinei , orient. II. Pontin de Mattos
Fortes, Renata, co-orient. III. Título.

Johana Maria Rosas Villena

Um método de apoio à autoria de vídeos acessíveis

Tese apresentada ao Instituto de Ciências Matemáticas e de Computação – ICMC-USP, como parte dos requisitos para obtenção do título de Doutora em Ciências – Ciências de Computação e Matemática Computacional. *VERSÃO REVISADA*

Área de Concentração: Ciências de Computação e Matemática Computacional

Orientador: Prof. Dr. Rudinei Goularte

Coorientadora: Profa. Dra. Renata Pontin de Mattos Fortes

USP – São Carlos
Junho de 2016

To the memory of my beloved family, Dora, Olga, Virginia, Rodolfo, Victor and Silverio. Their love and dedication to family taught me invaluable lessons in my life.

ACKNOWLEDGEMENTS

First, I would like to thank my family for their continued support and encouragement. Special thanks to my parents Jerry and Luz, and my aunts Vicky and Maritza. Thanks for our conversations, encouragement and interest in my thesis work. I would not be who I am today without you all. To my step-parents Eliseo and Aida thank you for being there. Thank you to my sisters and brothers, Vania, Leonardo, Ingrid, Dora, Valentina, and José for their support and partnership. I would like to thank my Peruvian family, for your support and encouragement.

I would like to thank my love, Luiz. You have been continually supportive all these years. Thank you for taking care of me all the time and for being my unconditional partner. Also, I would like to thank my Brazilian family. Thank you Bia and Paulo for taking care of me as part of your family and for all weekends together. Thank you Paulo, my brother in law, for sharing many moments, especially for your delicious cooking and traveling. Thank you Maré, Lucia, Joarez, Cecilia, Abílio, Zé Luiz, Suzana, João, Helena, Ligia, Asdrubal, Mariangela, Leda and Luiz Roberto for your amazing company.

I am very fortunate to have performed my doctorate work at a university as collaborative as the University of São Paulo; therefore, there are many people to thank for their part in my success. Specially, I would like to thank my advisors, Rudinei and Renata, for their guidance over the years. I am grateful for the opportunities they have afforded me, for our conversations, for their patience, kindness and generosity, and for their friendship. Under their mentorship I have learned the particulars of writing, which is an invaluable tool to have my career moving forward. Also, I would like to thank my brilliant students Bruno and Samuel for their collaboration in my research. Facilitas player and Transcript4All became true thanks to you.

I must acknowledge as well the many friends in the Intermedia Lab, Humberto, Tiago, Raoni, Kamila, Omar, Kishi, Kleber, João Paulo, Sandra, Olibário, Kyfayat, Raiza, Marcio, Flor, João, Luz, Bruna, John, Ana Luiza, Juliano, Marcio, Joab, and many others. Also, I would like to thank my São Carlos friends, Flávio, Igor, Domingos, Enrico, Conrado, Iván, Bruno, José, Eduardo, and Guilherme. I would like to thank my roommates Aninha, Liana, Bruna and Larissa for the great conviviality and friendship. Thanks to my longtime friends Jesús, Luz, Carla, and José for our friendship, I will always have you in my heart. Thank you Alessandro, Luciana, Americo, Clarissa, Akio, Bernardo and Sandra, I learned so much with you.

I would like to thank the program *Estudantes Convênio de Pós-Graduação (PEC-PG)*, of *CAPES/CNPq* for the financial support provided to this project.

Finally, I am grateful to many people who shared their experiences, especially to all volunteers of case studies for their time to contribute to this project. I would like to thank to the Educational Foundation of São Carlos (FESC) for collaborate with my research, specially to the teachers Adailton and Sandra, I have learned a lot of things thanks to you.

*“Imagination is everything.
It is the preview of life’s coming attractions.”
(Albert Einstein)*

RESUMO

ROSAS VILLENA, J. M.. **A method to support accessible video authoring**. 2016. 215 f. Doctoral dissertation (Doctorate Candidate Program in Computer Science and Computational Mathematics) – Instituto de Ciências Matemáticas e de Computação (ICMC/USP), São Carlos – SP.

As qualidades humanas somente são desenvolvidas na sociedade através da interação com os outros. Desde que a criança nasce, ela vê, ouve, toca, e sente as coisas. Portanto, estímulos visuais e auditivos são naturais para o homem, e estes estímulos também estão presentes em vídeos. Pode ser que seja esse o motivo pelo qual os vídeos causam grande impacto na sociedade. Uma evidência da popularidade e do crescimento de vídeos na sociedade é a grande quantidade de vídeos carregados no YouTube ou o grande número de vídeos postados no Facebook. Além disso, o surgimento dos smartphones tem aumentado a quantidade de usuários que não só assistem vídeos, mas também fazem vídeos. Mas o que acontece com as pessoas que não podem acessar estes conteúdos por terem alguma deficiência ou doença temporária? Para evitar que pessoas não compreendam o conteúdo do vídeo, é importante fornecer um vídeo com acessibilidade, assim, por exemplo, cerca de 23,9% do total da população de Brasil, declaradas com deficiência, serão beneficiadas. Diretrizes e leis para promover a acessibilidade para vídeos foram desenvolvidas. Existem também pesquisas sobre reprodutores de vídeo acessíveis, melhoramento da legenda, recursos sobre a língua de sinais e implementações para enriquecer conteúdos. Apesar do processo de produção de vídeo estar bem definido, as práticas e pesquisas atuais não consideram acessibilidade em seus estágios. Não se explica como autores amadores podem criar conteúdo alternativo considerando as diretrizes ou como incorporar essas diretrizes de acessibilidade no processo de produção de vídeo. Esta tese propõe um método, chamado Video4All, para a autoria de conteúdo alternativo dentro de um processo de produção de vídeo acessível. O Video4All é composto por um conjunto de atividades, incluindo diretrizes para conteúdo alternativo e avaliação desse conteúdo. Estudos de caso foram realizados para verificar a eficácia da aplicação do método proposto por autores amadores, avaliando o seu conteúdo alternativo criado. Assim, ao processo foi incorporada acessibilidade em todos os estágios, considerando-se estudos de caso que relataram as dificuldades que os usuários têm durante a interação com um reprodutor de vídeo acessível. Adicionalmente, foi desenvolvido um método de avaliação do conteúdo alternativo para obter uma medida de qualidade, de forma que os autores entendam melhor o nível de acessibilidade do conteúdo alternativo. O processo de produção de vídeo acessível pode ser utilizado por qualquer autor amador para melhorar a acessibilidade nos seus vídeos. O Video4All também auxilia na criação e avaliação de conteúdo alternativo.

Palavras-chave: autoria de conteúdo alternativo, processo de produção de vídeo acessível, acessibilidade para vídeos.

ABSTRACT

ROSAS VILLENA, J. M.. **A method to support accessible video authoring**. 2016. 215 f. Doctoral dissertation (Doctorate Candidate Program in Computer Science and Computational Mathematics) – Instituto de Ciências Matemáticas e de Computação (ICMC/USP), São Carlos – SP.

Human qualities are only developed in human society by means of interaction with others. Since a child is born, he/she sees, hears, touches, and tastes things. That means visual and aural stimuli are natural for human beings and these stimuli are also present on videos. Perhaps, this is the reason why videos have a huge impact on society. An evidence of the popularity of videos growing in society can be observed in the large volume of videos uploaded onto YouTube or the number of videos posted on Facebook. Besides, the rise of smartphones has increased the number of users that not only watch but also produce videos. However, what happens to people who cannot access those contents because they have disabilities or some temporary illness. To avoid excluding people from understanding video content, it is important to provide a video with accessibility. Thus, about 23.9% of the Brazil population who declared themselves with disabilities will benefit from accessible videos. Standards, laws and regulations to promote video accessibility have been developed. Research studies about accessible video players, improving captioning, resources about sign language, and implementation of the content enrichment have also been developed. In spite of the video production process being well defined, current research and market practices do not take accessibility into account as part of its stages. They do not explain how amateur authors can create alternative content following guidelines or how to incorporate accessibility guidelines to the video production process. This thesis proposes a method, called Video4All, for alternative content authoring as part of an accessible video production process. Video4All is composed by a set of activities, including alternative content guidelines, and alternative content evaluation. A case study was conducted to verify the effectiveness of applying the proposed method by amateur authors, evaluating their generated alternative content. In order to describe the process, was incorporated accessibility in all stages and the case studies reported the difficulties which users had during the interaction with an accessible video player. Additionally, the proposal involved the development of an alternative content evaluation method as a means to provide a quality measure, enabling authors to better understand the accessibility level of the alternative content they create. The accessible video production process can be used by any amateur author to improve the accessibility of their videos. Moreover, Video4All helps the authoring and evaluation of alternative content.

Keywords: alternative content authoring, accessible video production process, video accessibility.

LIST OF FIGURES

Figure 1 – Video production process	38
Figure 2 – Literature review focused on the post-production and distribution stages . . .	42
Figure 3 – An accessible video production process	52
Figure 4 – Facilitas Text Configuration	55
Figure 5 – Facilitas Search Control and Intentional Annotation Control	56
Figure 6 – Facilitas Caption Background Color and Link Control	60
Figure 7 – Facilitas Help Control	61
Figure 8 – Facilitas Controls Names	61
Figure 9 – New interface of Facilitas Player	64
Figure 10 – First Session with Dora	66
Figure 11 – Second Session with Dora	68
Figure 12 – Session with Alex	70
Figure 13 – Facilitas Player with the answer of the task.	76
Figure 14 – Video4All - Method for alternative content authoring	84
Figure 15 – ProDeaf Web Libras avatar	93
Figure 16 – Alternative content evaluation	95
Figure 17 – Facilitas - An accessible video player	101
Figure 18 – Case Study Planning	109
Figure 19 – Results from the alternative content activity using the proposed method . . .	115
Figure 20 – Results from the alternative content creation without using the method . . .	116
Figure 21 – Results from the alternative content creation using the method	117
Figure 22 – Author’s opinion about the information presented in the method	118
Figure 23 – Transcript template	173
Figure 24 – Transcript4All - Video information	176
Figure 25 – Transcript4All - Textual description	177
Figure 26 – Report of Ethic Committee	179
Figure 27 – Report of Ethic Committee	180
Figure 28 – Report of Ethic Committee	181
Figure 29 – Audio Description Coalition - Basics guidelines: Describe what you see . . .	204
Figure 30 – Audio Description Coalition - Basics guidelines: Describe objectively . . .	205
Figure 31 – Audio Description Coalition - Basics guidelines: Allow listeners to hear the dialogue	206

Figure 32 – Audio Description Coalition - Basics guidelines: Trust listeners’ ability to comprehend the material	207
Figure 33 – Audio Description Coalition - Basics guidelines: Censorship is unfair to the material and to listeners, and keep the language consistent	208
Figure 34 – Audio Description Coalition - Basics guidelines: Race, ethnicity and nationality	209
Figure 35 – Audio Description Coalition - Basics guidelines: Describe from listeners’ perspective, and good techniques make for good description	210
Figure 36 – Audio Description Coalition - Basics guidelines: Good techniques make for good description (continued from previous guideline)	211
Figure 37 – Audio Description Coalition - Standards unique to film and video description: Basics	212
Figure 38 – Audio Description Coalition - Standards unique to film and video description: Race, ethnicity and nationality, and techniques	213
Figure 39 – Audio Description Coalition - Standards unique to film and video description: Techniques (continued from previous guideline)	214
Figure 40 – Audio Description Coalition - Standards unique to film and video description: Techniques (continued from previous guideline)	215

LIST OF TABLES

Table 1 – Video Player Controls	31
Table 2 – Video players functionalities	49
Table 3 – Facilitas Player Versions	53
Table 4 – Features of videos	56
Table 5 – Participant’s Characteristics	57
Table 6 – Participant’s Characteristics	61
Table 7 – Questions for Captions evaluation	96
Table 8 – Questions for audio description evaluation	96
Table 9 – Questions for transcript evaluation	97
Table 10 – Example of requirements for caption	97
Table 11 – Quality measure of the accessibility level	99
Table 12 – Video’s characteristics	108
Table 13 – Author’s profile	109
Table 14 – Accessibility requirements evaluation for caption	112
Table 15 – Percentages of mandatory and recommended requirements covered by caption	113
Table 16 – Accessibility requirements achieved by authors	114
Table 17 – Results of the accessibility quality for each alternative content generated by authors	114
Table 18 – Requirements for caption	138
Table 19 – Requirements for audio description	149
Table 20 – Requirements for transcript	160
Table 21 – Author 1, video 1 (animation video)	164
Table 22 – Author 1, video 2 (documentary video)	165
Table 23 – Author 2, video 1 (documentary video)	166
Table 24 – Author 2, video 2 (animation video)	167
Table 25 – Author 3, video 1 (documentary video)	168
Table 26 – Author 3, video 2 (animation video)	169
Table 27 – Author 4, video 1 (animation video)	170
Table 28 – Author 4, video 2 (documentary video)	171

LIST OF ABBREVIATIONS AND ACRONYMS

ABNT	Brazilian National Standards Organization
AD	Audio description
ASL	American Sign Language
ATAG	Authoring Tool Accessibility Guidelines
CAAE	<i>Certificado de Apresentação para Apreciação Ética</i>
CC	Closed Caption
DCMP . . .	Described and Captioned Media Program
eMag	Electronic Government Accessibility Model
GQM	Goal-Question-Metric
IBGE	Brazilian Institute of Geography and Statistics
ISO	International Organization for Standardization
Libras	Brazilian Sign Language
OC	Open Caption
UAAG . . .	User Agent Accessibility Guidelines
UCD	User-Centered Design
VRS	Video Relay Service
W3C	World Wide Web Consortium
WAI	Web Accessibility Initiative
WCAG . . .	Web Content Accessibility Guidelines
WPM	Words Per Minute

CONTENTS

1	INTRODUCTION	23
1.1	The thesis	26
1.2	Organization	27
2	RELATED CONCEPTS AND LITERATURE REVIEW	29
2.1	Accessibility	29
2.2	Video production process	37
2.3	Literature Review	42
2.3.1	<i>Multimedia authoring</i>	44
2.3.2	<i>Accessibility in video content</i>	46
2.3.3	<i>Accessible video players</i>	47
2.4	Discussion	49
3	AN ACCESSIBLE VIDEO PRODUCTION PROCESS	51
3.1	Rationale for characterizing the process	52
3.1.1	<i>Iteration 1 - students and older people</i>	54
3.1.2	<i>Iteration 2 - adults and older people</i>	59
3.1.3	<i>Iteration 3 - users with auditory disabilities</i>	63
3.1.4	<i>Iteration 4 - blind user</i>	72
3.2	Pre-production	77
3.3	Production	80
3.4	Post-production	81
3.4.1	<i>Video editing</i>	81
3.4.2	<i>Video evaluation to identify seizure risks</i>	82
3.4.3	<i>Video4All - Method for alternative content authoring</i>	83
3.5	Distribution	99
3.5.1	<i>Facilitas - An accessible video player</i>	101
3.6	Discussion	104
4	METHOD VALIDATION	107
4.1	Planning	107
4.2	Collecting	109
4.3	Analysing and reporting	112
4.4	Discussion	119

5	CONCLUDING REMARKS	121
5.1	Answering the thesis questions	122
5.2	Main contributions	123
5.3	Limitations and future work	123
5.4	Results' Dissemination	125
	BIBLIOGRAPHY	127
	GLOSSARY	135
APPENDIX A	ACCESSIBILITY REQUIREMENTS FOR ALTERNA- TIVE CONTENT	137
APPENDIX B	VIDEO4ALL METHOD, CASE STUDY RESULTS . .	163
APPENDIX C	TRANSCRIPT TEMPLATE	173
APPENDIX D	TRANSCRIPT4ALL TOOL	175
ANNEX A	REPORT OF ETHIC COMMITTEE	179
ANNEX B	GUIDELINES FOR CLOSED CAPTION	183
B.1	Text	183
B.2	Language Mechanics	187
B.3	Presentation Rate	189
B.4	Sound Effects and Music	192
B.5	Speaker Identification	194
B.6	Special Considerations	195
B.7	Appendices	197
ANNEX C	STANDARDS FOR AUDIO DESCRIPTION	203

INTRODUCTION

Sociability is humans' natural instinct and every human quality such as thinking, enquiring, learning languages, playing and working are only developed in human society by means of interaction with others. From birth to death, humans acquire different social skills by social interaction with their fellow beings which mould their personality (HOSSAIN; ALI, 2014). When a child is born, it sees, it hears, it touches, and it tastes things. So that "seeing comes before words. The child looks and recognizes before it can speak" (BERGER, 2008). Thus, visual and audio stimuli are natural for humans and these stimuli are also present in videos, which may be one reason why videos have a huge impact on society.

The interest of society in watching things can be observed in the popularity of YouTube¹ videos, Netflix², TED³, Snapchat⁴, Instagram⁵, and Facebook⁶.

Taking YouTube as an example, more than 300 hours of new videos are upload every minute, and every day people watch hundreds of millions of hours on YouTube and generate billions of views (YOUTUBE, 2015). Another evidence of videos' growth is that the number of video posts per person on Facebook has increased 75% globally since 2014 (FACEBOOK, 2015). Furthermore, according to Forbes, one of the top seven online marketing trends that will dominate 2016 is the growth of video advertisements due to users are becoming more accepting of video ads online (DEMERS, 2015). It is predicted that this trend will continue in the future. Cisco explain that globally consumer Internet video traffic will be 80% percent of all consumer Internet traffic in 2019, up from 64% in 2014 (CISCO, 2014).

¹ YouTube is the most popular online-video sharing Web site - <<https://www.youtube.com/>>

² Netflix is a global provider of streaming movies and TV series - <www.netflix.com>

³ TED is a nonprofit website devoted to ideas worth spreading which usually shows talks between 10 and 20 minutes - <www.ted.com>

⁴ Snapchat is a video messaging application that allows posting videos and images - <www.snapchat.com>

⁵ Instagram is a social network that allows photo and video sharing - <www.instagram.com>

⁶ Facebook is a social network that among all the features allow posting photos and sharing videos - <www.facebook.com>

The growth of online videos could be influenced by the increase in Internet speed, availability of higher screen resolutions, the possibility to create responsive web pages, new video compression types and techniques that can display on any screen size from cellphones to movie screens (REA; IRVING, 2015).

The rise of the cellphone has increased youngsters' and adults' interest in not only watching videos but also making videos. Surveys conducted by Pew Research Center's Internet Project show an overview of this. In 2013, 54% of adult Internet users posted original photos or videos online that they themselves had created (CENTER, 2013). In 2014, the video consumption is part of everyday life of youngsters and adults in the United States. Youngsters were the greatest consumers of online videos, nine out of ten 18-to-29-year-olds watching online videos and more than six out of ten adults watching videos online (CENTER, 2014). In 2015, 89% of cellphone owners stated they used their phone during the most recent social gathering they attended and 70% of smartphone owners took a photo or video during their most recent social gathering.

Until now, it was discussed that seeing, hearing, touching and tasting things are natural to any individual in a society and videos are a way to interact with other peers. However, what happens to people who cannot see, hear, touch or taste things because they are people with disabilities or because they have a temporary impairment? To avoid excluding those people from understanding videos content, it is important to provide alternatives to understand video content and allow access to it, independently of disabilities. In other words, it is important to provide accessible videos.

To be aware of the large quantity of people who benefit from with video accessibility, the following data were observed. Over a billion people, about 15% of the world's population, have some kind of disability (ORGANIZATION, 2015). In the 2010 Census conducted by the Brazilian Institute of Geography and Statistics (IBGE), 23.9% of the Brazilian population had declared themselves as having a disability (SOCIAL, 2012). Also in 2010, according to the Census Bureau, 19% of the population of United States had a disability (BRAULT, 2012).

Looking for clues about how video accessibility is being approached, this study examined the top 50 most accessed websites according to <www.Alexa.com> (ROSAS-VILLENA *et al.*, 2015). From those 50 websites, 34 were excluded from further analysis because they did not show any video or because the language was different from English, Spanish or Portuguese. Other 6 sites (Twitter, LinkedIn, Blogspot, Hao123, Wordpress and Reddit) were also excluded because they show videos hosted in other sites like YouTube or Vimeo. The 10 remaining websites were selected - they allow watching or uploading videos (Facebook, YouTube, Yahoo, Vk, Msn, Instagram, Tumblr, Apple, Xvideos and Microsoft). All these websites present play and pause controls. Three of them (Instagram, Tumblr and Microsoft) do not have volume control. Only one site (Microsoft) does not have screen size configuration. Currently, Microsoft only has two basic functionalities: play and pause. There are four sites (YouTube, Vk, Msn, Xvideos) which have quality configuration. Only two websites (YouTube and Yahoo) present caption configurations,

even though few videos present captions. The keyboard navigation in some sites is irregular. However, these controls do not guarantee that the video player is accessible, they help people with disabilities understand the video content (through controls for alternative content such as closed caption, audio description, transcript and sign language), change their preferences (caption settings such as caption color, size, font and background color), and find content easily (search, links, chapters, forward/rewind, speed down and speed up). When the video player provides options to adding alternative content, authors are motivated to create alternative content.

Standards, laws and regulations to promote video accessibility were developed. Examples include recommendations for web content described in Web Content Accessibility Guidelines (WCAG) 2.0, guidelines for designing user agents to help disabled people are described in User Agent Accessibility Guidelines (UAAG) 2.0, Section 508 in the USA and the Electronic Government Accessibility Model (eMag) 3.1 in Brazil. Also, research studies about video accessibility have been developed in the last years. For example, studies about the development of accessible video players (CHENG *et al.*, 2009), (LI *et al.*, 2009), (GONZÁLEZ *et al.*, 2011), (ABLEPLAYER, 2015), (AFB-ACCESSIBLEPLAYER, 2015), improving captioning (HONG *et al.*, 2011), resources about sign language (ANTUNES *et al.*, 2011), (LOPEZ-COLINO; COLAS, 2012), (ARAÚJO *et al.*, 2014), and implementation of content enrichment (ENCELLE *et al.*, 2011), (CHAMPIN *et al.*, 2010), (RODRIGUES *et al.*, 2010), (VILLAMIZAR *et al.*, 2011). These studies developed ways to help people with disabilities to understand the content; however, none of them explains how to create alternative content following guidelines for amateur authors or how to incorporate accessibility guidelines in the video production process. When it concerns incorporating accessibility in video content, some guides were found, some focused in creation of educational material (CARRERA *et al.*, 2013), (HILERA-GONZÁLEZ; CAMPO-MONTALVO, 2015), and one other focused for web developers (PAE, 2014).

Although the video production process is well defined, since the same stages (pre-production, production, post-production, and distribution) did not changed in ten years as it is observed by Compesi and Gomez (2005) and Rea and Irving (2015), they did not incorporate accessibility into the process. On the other hand, research studies about multimedia authoring are concentrated on the creation of models, templates, multimedia presentations, among others. Researchers usually explain different authoring paradigms and also discuss implementation languages. Rowe (2013) explained that authoring tools for multimedia content improved between 2003 and 2013. It is relatively easy to capture images, audio, and video using smartphones and incorporate them into applications. The development of HTML5, in particular regarding tags, is likely to lead to further improvements in supporting continuous media. Also, Rowe (2013) explained that there are other areas of authoring in progress as the recognition of different types of video content and the development of templates for authoring and/or consuming the content, the creation and use of lecture videos for learning, and the support for linking different elements with additional material and supporting interactivity are still difficult to the author.

Thus, two main necessities emerge. The first one is related to defining an accessible video production process and the second is about defining a method for alternative content authoring. These two requirements were the main issues of investigation of this thesis.

Regarding the first requirement, the video production process consists of four stages, in which activities and recommendations to capture the video are defined. Incorporating accessibility to this process from the first stage helps the final product to be more accessible than if accessibility had not been incorporated. To describe the process, this study considered case studies that reported the difficulties that users had during the interaction with an accessible video player.

The second requirement involves the accessibility guidelines related to video content which explain some recommendations and tools for alternative content, but did not explain in details those recommendations or how to create alternative content step by step, neither how to validate the alternative content created. Because of that, it is necessary to create a method that connects all these elements, including the incorporation of alternative content as text (transcript and captions), audio (audio description) or video (sign language). This method should allow to be followed by any amateur author.

1.1 The thesis

Given: (i) the benefits of accessibility guidelines; (ii) the value of a method for alternative content authoring; and (iii) the necessity of accessible video production process; the main research question proposed on this thesis is:

Can the accessibility guidelines and an accessible video production process that include them, be combined in a method to be able to generate video with better accessibility levels as result?

The thesis is: the video accessibility level can be increased by using a method for alternative content authoring as part of an accessible video production process. In this work, the thesis is an individual piece of work pursued emphasizing accessibility issues during all stages of the process, including pre-production, production, post-production and distribution.

In order to validate this thesis, the proposal incorporated a method for alternative content authoring, composed by activities with guidelines and metrics for alternative content evaluation inside of an accessible video production process.

In order to achieve that goal, the following four research questions were pursued:

1. Which elements are necessary for a process to support accessible video authoring capable of promoting accessibility?
2. How can a method increase the accessibility level of a video?

3. How can a method be included as part of the accessible video authoring process?
4. How to evaluate whether the levels of accessibility of the alternative content are compliant with the guidelines established?

1.2 Organization

This thesis is organized as follows:

Chapter 2 presents the related concepts about video accessibility, alternative content guidelines, and literature review about multimedia authoring, accessibility in video content and accessible video players.

In Chapter 3 preliminary studies for characterizing the process and the process to support accessible video authoring are described. This process was adapted to observe the interaction with an accessible video player, called Facilitas (see [subsection 3.5.1](#)), during four case studies (see [section 3.1](#)). Also, this chapter describes a method (called Video4All) for alternative content authoring, which incorporates activities as: guidelines, activities to address the authoring process, tools to help with the authoring activities and describe metrics for evaluation of each alternative content.

Chapter 4 presents a case study conducted with video authors in order to evaluate whether the levels of accessibility using the method were the same or better than without using it. This study highlights the main contributions and results of this research.

Finally, Chapter 5 presents the concluding remarks and the future works regarding the proposal of this thesis.

RELATED CONCEPTS AND LITERATURE REVIEW

This chapter presents two important topics towards the understanding of this thesis: related concepts and literature review. First, concepts regarding accessibility, video accessibility and alternative content guidelines are presented at Section [section 2.1](#), providing the reader with the basics about accessibility in the Computer Science field. Then, at [section 2.2](#), the video production process is discussed revealing a lack for accessibility. Finally, [section 2.3](#) presents a literature review, discussing current research about video accessibility and pointing out some gaps.

2.1 Accessibility

A widely-accepted definition for the term accessibility refers to the extent to which products, systems, services, environments and facilities can be used by people from a population with the widest range of characteristics and capabilities to achieve a specified goal in a specified context of use ([ISO/IEC71, 2014](#)).

“Web accessibility means that people with disabilities can perceive, understand, navigate, and interact with the Web, and that they can contribute to the Web. Web accessibility also benefits others, including older people with changing abilities due to aging” ([W3C, 2012](#)).

Specific guidelines for web accessibility were developed by World Wide Web Consortium (W3C), and later endorsed by the International Organization for Standardization (ISO). The W3C Web Accessibility Initiative (WAI) has standardized a set of guidelines for the accessibility content. Recommendations for web content are described in the Web Content Accessibility Guidelines (WCAG) 2.0 ([W3C, 2008e](#)), and guidelines for design user agents (browsers, video players, plugins) to help people with disabilities are described in the User Agent Accessibility Guidelines (UAAG) 2.0 ([W3C, 2015d](#)). Guidelines that define how authoring tools should

help web developers produce web content are accessible in the Authoring Tool Accessibility Guidelines (ATAG) 2.0 (W3C, 2015b).

WCAG 2.0 is a stable and referable technical standard. It has 12 guidelines that are organized under 4 principles: perceivable, operable, understandable, and robust. For each guideline, there are testable success criterion, which are at three levels: A, AA, and AAA (W3C, 2008e).

User Agent Accessibility Guidelines (UAAG) 1.0 was approved in December 2002 and it is the stable and referable version. Currently, UAAG 2.0 is a mature draft in which the main content will not change significantly. In this thesis, the study performed considered UAAG 2.0 W3C Working Draft updated in September, 2015 (W3C, 2015d).

The Authoring Tool Accessibility Guidelines (ATAG) 2.0 provides guidelines for designing web content authoring tools that are both more accessible to authors with disabilities (Part A) and designed to enable, support, and promote the production of more accessible web content by all authors (Part B) (W3C, 2015b).

The ISO is responsible for international management standards. Among them, the ISO 9241-171:2008 provides ergonomic guidance and specifications for the design of accessible software (ISO9241-171, 2008). It covers issues associated with the design of accessible software for people with physical, sensory and cognitive abilities, including those temporarily disabled and the elderly. The ISO/IEC GUIDE 71:2014 provides guidelines to addressing accessibility in standards and explain the human abilities and characteristics (ISO/IEC71, 2014).

Laws and regulations have been defined to protect the user's rights. The law of information and communication access N^o 13.146 of July 6th, 2015 of Brazil in the Article 63, establishes that accessibility on websites is mandatory by commercial companies and government agencies (LAW13.146, 2015). Examples of specific governmental recommendations for web accessibility include the eMag 3.1 model in Brazil (EMAG, 2014), Section 508 in the USA (SECTION508, 2000), and the European Directive Accessibility requirements suitable for public procurement of ICT products and services in Europe (CEN/CENELEC/ETSI, 2014).

Video accessibility

Video, as a multimedia element, should be accessible by people from a population with the widest range of characteristics and capabilities to understand the video content. An accessible video should provide a synchronized alternative content (at least closed caption and audio description or media alternative for prerecorded video) and must be reproduced in synchronization of a user agent. User agents (browsers, video players, etc.) must allow access to video content so that any user can interact in an intuitive and effective way.

Following are described initiatives that provide specific guidelines to video accessibility:

- Video validation to identify seizure risks in a common guideline cited by WCAG 2.0,

Table 1 – Video Player Controls

Name	Description	UAAG 2.0	ISO 9241-171
Play	Play the video content	2.10.5	10.8.2
Pause	Pause the video content	2.10.5	10.8.2
Stop	Stop the video content	2.10.5	10.8.2
Resize	Resize the viewports	1.8.3	10.5.8
Mute	Enable or disable the audio content	1.5.1	10.6.2
Volume	Adjust the volume	1.5.1	10.6.2
Forward	Forward seconds within a reproduction	2.10.6	10.8.3
Rewind	Delay seconds within a reproduction	2.10.6	10.8.3
Speed down	Speed down the video and audio content	2.10.4	
Speed up	Speed up the video and audio content	2.10.4	
Caption	Enable or disable captions	1.1.1, 1.1.2	10.1.3, 10.7.2
Audio description	Enable or disable audio description	1.1.1, 1.1.2	10.1.3
Sign language	Enable or disable sign language	1.1.1, 1.1.2	-
Transcript	Enable or disable transcript	1.1.1, 1.1.2	-
Color	Change the color of the captions	1.4.1	10.7.3
Font-size	Change the size of the captions	1.4.1	10.7.3
Font-Family	Change the font of the captions	1.4.1	10.7.3
Background	Change the background of the captions	1.4.1	-
Line spacing	Change the line spacing of the captions	1.4.1	-
Search	Search within captions	2.4.1, 2.4.5	-
Search result	When a search operation produces a match, it is showed a list	2.4.3	-
Alert for no match	The user can choose to receive notification when there is no match to a search operation	2.4.4	-
Help	Documentation about accessibility features	3.2.1	11.1.5
Controls names	Provide controls names	3.2.1, 1.10.1	-
Language Caption	Change the language of the captions	2.6.1	8.2.1
Language Audio	Change the language of the audio description	2.6.1	8.2.1
Annotations or Chapters	Navigate by annotations, chapters or scenes	2.10.7	-
Keyboard navigation	Provide Full Keyboard navigation	2.1.1	-
Keyboard Focus	Every viewport has an active or inactive keyboard focus at all times	2.1.2	-

UAAG 1.0 and 2.0, and ISO 9241-171. Guideline 2.3 of WCAG 2.0 refers to “do not design content in a way that is known to cause seizures”, this guideline contains two success criterion: “three Flashes or Below Threshold” (Level A) and “Three Flashes” (level AAA).

Guideline 3.3 of UAAG 1.0 refers to “allows configuration not to render some content that may reduce accessibility - Toggle animated or blinking text”. Guideline 2.9 of UAAG 2.0 refers to “helps users avoid flashing that could cause seizures”, this guideline includes the success criterion “three Flashes or Below Threshold” has Level A. Finally, guideline 10.1.1 of ISO 9241-171 refers to “avoids seizure-inducing flash rates”.

- The WCAG 2.0 also presents the guideline 1.2 “Time-based Media: it provides alternatives for time-based media” (W3C, 2008e), in which is specified the addition of alternative content as media alternative (can be transcript), caption, audio description, and sign language. According the levels of conformance, the accessibility requirements of a prerecorded video must comply with: Level A (media alternative or audio description, caption and seizure validation), Level AA (audio description) and Level AAA (Sign language, extended audio description and media alternative). To assign one of the three levels of conformance, the working group should take into account a wide range of interacting issues. To assign level AA, it is necessary comply level A. For example, considering that a video passed the seizure validation, if that video has audio description and caption, it can be considered level AA, but if that video have a media alternative (transcript) and caption, it can be considered level A.
- The UAAG 2.0 presents guidelines for media players, for example, Guideline 1.1: Alternative content, Guideline 1.5: Volume configuration or Guideline 2.10 - Provide control of time-based media (W3C, 2015d).
- The ISO 9241-171:2008 also provides guidelines for media players, for example, 10.7 Text equivalents of audio (captions), 10.8 Media and animation or 11.1 Documentation and Help (ISO9241-171, 2008).
- The 21st Century Communications and Video Accessibility act signed in the USA in October, 2010 promotes expanded access to Internet-based video programming (COMMISSION, 2010), for example, “Requires video programming that is closed captioned on TV to be closed captioned when distributed on the Internet (does not cover programs shown only on the Internet)”.

Table 1 shows a list with controls that an accessible video player could provide, success criterion from UAAG 2.0, and guidelines from ISO 9241-171.

Alternative content guidelines

Alternative content fulfills essentially the same function or purpose as the original content (W3C, 2015a). There are several general types of alternative content: text alternative for non-text content (images), alternative for time-based media (audio and video), text alternatives for text content (abbreviation) and media alternative for text. As the focus of this research is videos,

the alternative for time-based media should be explained. Alternative for time-based media is a Web content that serves the same function or purpose as one or more tracks in a time-based media presentation (W3C, 2015a). This includes alternatives for audio such as captions and sign language interpretation, and alternatives for video such as audio descriptions and extended audio descriptions. Another form of time-based media alternative is a correctly sequenced text description of time-based visual and auditory information which is also capable of achieving the outcomes of any interactivity in the time-based presentation.

In short, the most common types of alternative content for videos are: caption, audio description, text description (a transcript include text description) and sign language. Captions are the text alternative to speech and non-speech audio information (W3C, 2008b). Audio description is a narration added to the soundtrack to describe important visual details that cannot be understood from the main soundtrack by itself (W3C, 2008a). Transcripts are the text of what was said in the video and relevant visual and non-speech audio information; it does not require any time codes (W3C, 2008a).. Sign language is a language using combinations of movements of the hands and arms, facial expressions, or body positions to convey meaning (W3C, 2008c). The following sections describe each alternative content.

Captions

Captions are alternative text to media, which provide the part of the content available via the audio track, and they are synchronized with the audio track. Captions are not needed when the synchronized media is, itself, an alternate presentation of information that is also presented via text on the Web page (DIGITALGOV, 2014a; W3C, 2008e).

Captioning is necessary for people who are unable to hear the audio in a video. It's also helpful for people with cognitive impairments, as well as for developing literacy, both in children and adults (DIGITALGOV, 2014a).

There are two types of caption:

1. Open Caption (OC). Which is always visible, the words appear automatically on the video when it is played; it can not be turned off.
2. Closed Caption (CC). Which only appear when it is turned on and it can be turned off. It includes not only dialogue, but also non-speech information such as speaker identification, music, laughter and meaningful sound effects.

Captions are different from subtitles. Captions are usually in the same language as the spoken, while subtitles usually translate text into another language, and are common in films (DIGITALGOV, 2014a). Subtitles provide only the content of spoken dialogue and captions add equivalents for non-dialogue audio information (W3C, 2008e).

Following are described the guidelines to provide captions for audio content in a synchronized media:

- The WCAG 2.0 success criterion 1.2.2 Captions (Prerecorded) of the guideline 1.2 Time-based Media: Provide alternatives for time-based media.
- The checkpoint 2.5 Make captions, transcripts, audio descriptions available of the guideline 2: Ensure user's access to all content of the UAAG 1.0.
- Guideline 1.1 - Provide access to alternative content of UAAG 2.0.
- The eMag 3.1, Electronic Government Accessibility Model of Brazil, also recommends the use of captions (EMAG, 2014).
- The Section 508 Standards for Electronic and Information Technology has the standard 1194.24 - Video and multimedia products indicate that closed captions are necessary for the comprehension of the content (SECTION508, 2000).

Audio description

Audio description (AD), also called descriptive video, video description or descriptive narration, is an additional audio track to describe important visual details that cannot be understood from the main soundtrack by itself. AD provides information about actions, characters, scene changes, on-screen text, and other visual content. The narration is added during existing pauses in dialogue (DIGITALGOV, 2014b; W3C, 2008e).

AD makes videos and multimedia accessible to people who have "low vision" (very poor vision), who are blind (by capturing what is happening on screen into audible descriptions that are played during natural pauses in the audio track), who have cognitive disabilities (difficulties to interpret visual information or people with learning disabilities, for example, students that learn a new language, vocabulary acquisition, association between image and word).

There are three types of AD (DIGITALGOV, 2014b):

1. Closed AD. These descriptions can be turned on or off by the viewer.
2. Open AD. These play automatically when the video is played and cannot be turned off.
3. Real-time AD. This type of description provides live commentary or narration by a trained audio describer via headphones and a small transmitter. It applies to theatrical productions, in-person tours, and other live events, as opposed to pre-recorded videos.

Following are described the guidelines to provide AD in a synchronized media:

- The guideline 1.2 Time-based Media: Provide alternatives for time-based media of WCAG 2.0, provide three success criterion related to audio description: 1.2.3 Audio Description or Media Alternative (Prerecorded), 1.2.5 Audio Description (Prerecorded) and 1.2.7 Extended Audio Description (Prerecorded).
- The guideline 2: Ensure user access to all content of UAAG 1.0 and it provides the checkpoint 2.5 Make captions, transcripts, audio descriptions available.
- Guideline 1.1: Provide access to alternative content of UAAG 2.0.
- The eMAG, also recommend the use of audio description ([EMAG, 2014](#)).
- The Section 508 Standards for Electronic and Information Technology has the standard 1194.24 - Video and multimedia products indicate that AD need to be selected by the user ([SECTION508, 2000](#)).

Transcript

Transcript is an alternative content for synchronized media. Its purpose is to provide a description of the video in text to satisfy user's needs. Full descriptions are provided of all visual information, including visual context, actions and expressions of actors, and any other visual material. In addition, non-speech sounds (laughter, off-screen voices, etc.) are described, and transcripts of all dialogue are included. The sequence of description and dialogue transcripts are the same as the sequence in the synchronized media itself ([W3C, 2008a](#)).

A full text transcript should include information that would be in both the caption and video description, so that it is a complete representation of the material, as well as containing any interactive options ([W3C, 2015c](#)).

Transcript benefits:

- People with auditory disabilities ([ABOU-ZAHRA, 2012](#); [DO-IT, 2013](#)).
- People who are unable to see the content and who therefore would be unable to access captions (they can read the transcript using assistive technologies such as Braille output devices) ([DO-IT, 2013](#)).
- People who will not spend the time to listen to the audio or watch the video, but will skim a transcript ([LAWTON, 2009](#)).
- People who have cognitive disability, they have difficulty to processing auditory information or understand the video content ([LAWTON, 2009](#); [DIGITALGOV, 2013, 2013](#)).
- People who are not proficient in the language who find it easier to read than listen or they may struggle with unfamiliar words ([LAWTON, 2009](#); [DIGITALGOV, 2013](#)).

- People who are non-technical users, if the video contains jargon ([DIGITALGOV, 2013](#)).
- People with low bandwidth connections ([DO-IT, 2013](#)) who do not want to download the larger audio or video file ([LAWTON, 2009](#)).
- People who pay for bandwidth usage and thus do not want to download the larger audio or video file. This is often an issue with phones and other mobile devices ([LAWTON, 2009](#)).
- People who cannot play the audio because they are in a noisy environment and they cannot hear it ([LAWTON, 2009](#)).
- People who cannot play the audio because they are in a quiet environment and they do not want to disturb others ([LAWTON, 2009](#)).
- Promote more traffic to the video, transcript help with search engine optimization. Search engines can index the transcript, not the audio or video ([LAWTON, 2009](#); [DIGITALGOV, 2013](#)).

Following are described the accessibility guidelines to provide transcript:

- The guideline 1.2 Time-based Media: Provide alternatives for time-based media of WCAG 2.0, provide two success criterion 1.2.3 Audio Description or Media Alternative (Prerecorded) with level A and 1.2.8 Media Alternative (Prerecorded) with level AAA.
- The success criterion 1.2.1: Audio-only and Video-only (Prerecorded) of WCAG 2.0 explains that alternative for time-based media (transcript) or audio track is necessary for videos without sound to provide an equivalent to what is presented visually. If the video has no sound, inform about it to users to avoid discomfort and they can concentrate on the video without having to look for alternative content.
- The guideline 2: Ensure user access to all content of UAAG 1.0 provide the checkpoint 2.5 Make captions, transcripts, audio descriptions available.
- The guideline 1.1: Provide access to alternative content of UAAG 2.0 with the success criterion 1.1.5 provide Configurable Alternative Content Defaults.
- The eMAG also recommend the use of transcript ([EMAG, 2014](#)).

Sign Language

Sign language enables people who are deaf or hard of hearing and who are fluent in a sign language to understand the content of the audio track of videos. Written text, such as that found in captions, is often a second language to them. Because sign language provides the ability to provide intonation, emotion and other audio information that is reflected in sign

language interpretation, but not in captions, sign language interpretation provides richer and more equivalent access to synchronized media (W3C, 2008c).

People whose human language is a sign language sometimes have limited reading ability. These individuals may not be able to read and comprehend the captions and thus require a sign language interpretation to gain access to the synchronized media content (W3C, 2008c).

In Brazil, according to the census 2010 of IBGE, there are 9,7 millions of people with hearing disabilities (SOCIAL, 2012). *Brazilian Sign Language (Libras)* is the Brazilian second official language, established by the law 10.436/2002 and regulated by decrees 5.296/2004 and 5.626/2005, which requires the application of *Libras* for communication and accessibility to deaf people in the country.

Below are described the guidelines to provide sign language for prerecorded audio content in synchronized media.

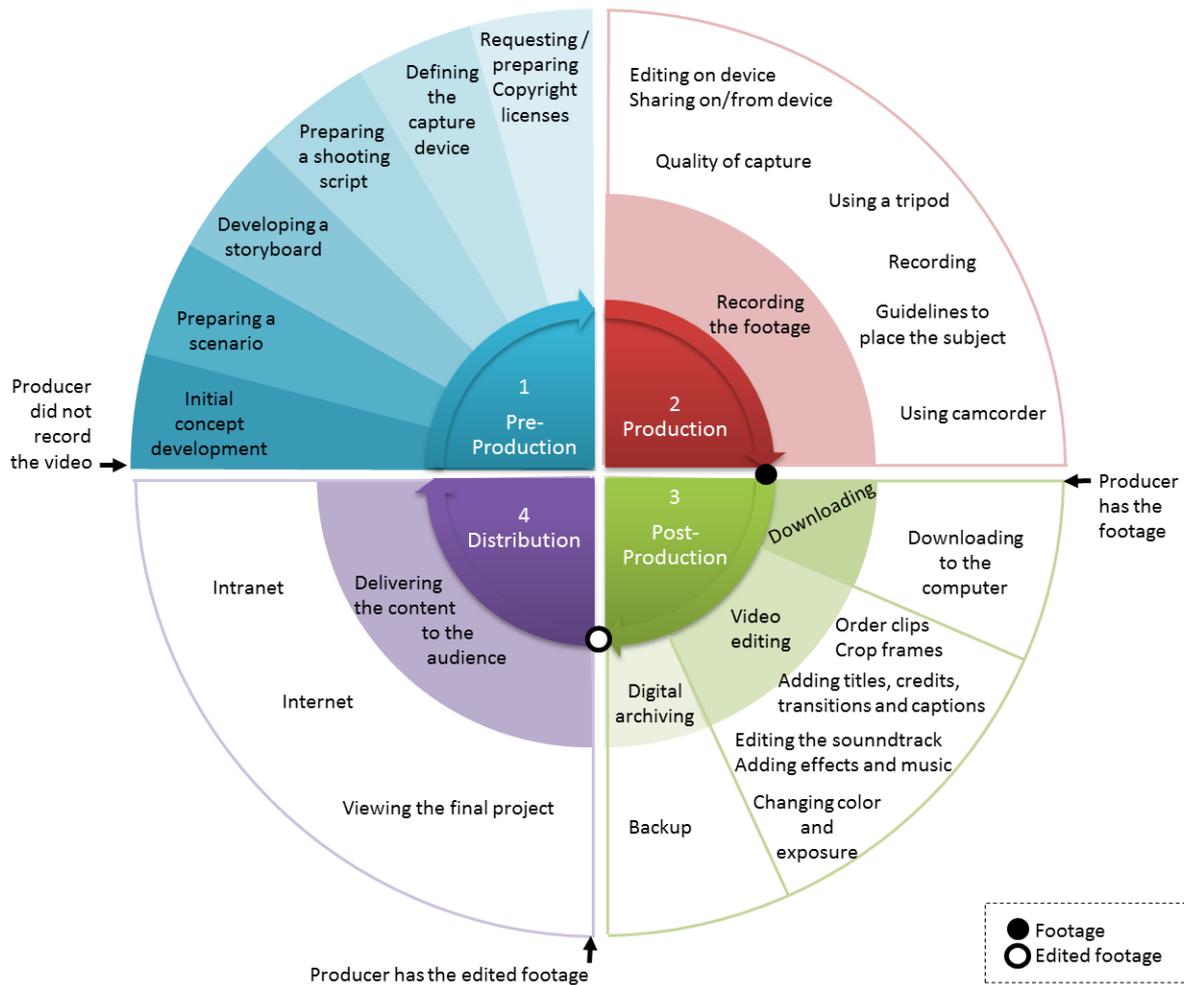
- The success criterion 1.2.6: Sign Language (Prerecorded) with level AAA of the guideline 1.2 Time-based Media: Provides alternatives for time-based media of WCAG 2.0.
- The success criterion 1.1.4 Facilitate Clear Display of Alternative Content for Time-based Media, 1.1.6 Use Configurable Text for Time-based Media Captions and 1.1.7 Allow Resize and Reposition of Time-based Media Alternatives of the guideline 1.1: Provide access to alternative content of UAAG 2.0.
- The eMAG also recommends the use of sign language (EMAG, 2014).

2.2 Video production process

The video production process refers to the stages required to develop a video, from the idea to the final product. All video production process involves four stages: (1) pre-production, (2) production, (3) post-production, and (4) distribution. Each stage is composed by activities. Each activity has its own challenges and it will vary depending on the size of the project, the amount of time, effort and the investment value. Based on Carrera *et al.* (2013), Media and Education (2012), Compesi and Gomez (2005), and Kirk *et al.* (2007), the video production process was defined for the purposes of this research. Figure 1 shows the four stages and its activities that compose the video production process.

The professional in charge of the control and completion of a project is the producer. The producer is responsible for developing the program idea, hiring the production crew to execute the production and making sure that the project is completed within budget and on time. The following sections describe the stages of pre-production, production, post-production, and distribution.

Figure 1 – Video production process



Source: Elaborated by the author.

Pre-production

The pre-production stage is where all the planning takes place before the production begins. Carrera *et al.* (2013), Media and Education (2012) and Kirk *et al.* (2007) reported activities for the pre-production. These activities are summarized as follows:

- **Initial concept development.** Define the topic or idea, and determine the intended audience.
- **Preparing a scenario.** Identify how long the program is to be, what type of approach you will be taking to the topic, the overall structure that will be used and the basic types of scenes that will make up the video. The scenario is often a critical document because it can be used to educate others about the project.
- **Developing a storyboard.** The idea is broken into small segments, each segment contains visual representations of what is to be seen, the proposed written script and any instruction

such as angles of the camera, music background, and desired volume.

- **Preparing a shooting script.** Provide detailed directions to the camera operator on what to shoot and how to shoot it, and the specific information about what the audio track is supposed to contain. It also includes details on any graphics required and how they are to be integrated into the video.
- **Defining the capture device.** Define the video device to be used in the production stage. This device can be an specialized device (camcorder) or any device that allow recording (digital cameras, cellphones, etc.). According the importance of an event, individuals identify the type of device to be used, for example, they select specialized devices to celebrate special events that were planned, or select any device to record activities that were not planned.
- **Requesting copyright licenses.** Request licenses for the participation of people, institutions and entities. For children or people with disabilities, it is required the license of parents or the legally responsible person. For third party videos, it is also request the license, except when the video is didactic, it could be permitted by authority according to the country's law.
- **Preparing copyright licenses.** Prepare the necessary licenses for the recordings.

Production

The production stage refers to the part of the process in which footage is recorded. The goal is to record all required shots. Shooting may extend over days or weeks, depending on the length of the project (COMPESI; GOMEZ, 2005).

According to Kirk *et al.* (2007), the production stage can be divided into capture, edit on device, and share on/from device. These activities are detailed as follows:

- **Capture.** Video capture is the process of converting an analog video signal to digital video. They define two schemas of video capture, spontaneous and ad-hoc capture. Spontaneous capture usually is associated with any device that allow recording. Ad-hoc capture is much more planned, with attention given to the quality of the device.
- **Edit on device.** Editing at the time of capture is usually frequent on spontaneous capture. While for ad-hoc capture it happens later on the post-production stage.
- **Share on/from device.** Showing the captured video at the time of taking it is also frequent on spontaneous capture and for ad-hoc capture it happens after editing the content.

Kirk *et al.* (2007) also remarked the quality of capture needs to be adjusted in situ, before it is too late. The quality of capture is related to display resolution. Display resolution is the

number of distinct pixels in each dimension that can be displayed, it is usually quoted as width x height. For example, 1280 x 720 (HD 720), 1920 x 1080 (HD 1080). The producer should select the best resolution that best fits the project needs.

During the production, [Media and Education \(2012\)](#) and [Carrera *et al.* \(2013\)](#) recommended some orientations to know how the professionals work with camera placement, angle, point of view, selection of backgrounds, use of colour, and balance when they are shooting. These recommendations are explained as follows:

- **Learning how to use the camcorder.** Learn the locations of all of the basic controls on the camcorder, for example, where the back light control is and how much adjustment it can provide, how to switch from manual to auto focus and where the macro control is for close-up shooting, and how to perform special effects in the camera.
- **Recording start-up and ending.** Start recording ten to fifteen seconds in advance of the beginning of the activity to ensure that the video signal is stable by the time the action begins. Also when the scene ends, record an extra ten to fifteen seconds to make sure that any rollback which can occur when the camera is turned off will not over top of something that was recorded.
- **Using a tripod.** Video tripods have been designed to help to eliminate annoying jerkiness commons in many amateur video productions and to provide controlled movement in both horizontal and vertical directions.
- **Learning about light.** The environment has many different types of light sources and these lights have different set of colour characteristics, for example, daylight is towards the blue end of the scale, fluorescents include a lot of green and household lighting is on the yellow end of the scale.
- **Understanding how the audio system in the camcorder works.** Use an external microphone to capture the surround sound.
- **Studying where a subject should be placed within the screen.** There are guidelines that help to understand how to do this.
 1. Rule of thirds. The first step is to divide the camcorder screen into thirds horizontally and vertically so that it has nine even squares. In order to create more visual interest, place the subjects if they are horizontal along the horizontal lines or if they are vertical along the vertical lines. Place a person or object at one of the junction points of the horizontal and vertical lines. This creates a far more interesting composition than one where the subject is right in the middle of the screen.
 2. Rule of 180°. When two people talk, record all of the shots on the same side of the subjects.

3. Lead space for movement. Provide some space or lead room in front of the subject for them to move into.
4. Balance and tension. Another aspect of composition has to do with the amount of visual tension or balance when setup the shots. If two people of the same height are placed so that they are both lined up along the vertical lines of the rule of thirds layout that would be a very balanced scene with almost no visual tension. On the other hand, if there is a person lined up in the bottom third of one of the vertical lines and a tall object, like a building crane, which extends from the bottom to the top of the frame along the opposite side, it has created a dynamic tension in the scene between these two visual elements.

Post-production

The post-production stage begins after all the footage has been captured. This stage consists of bringing together all of the material created in the production stage to form the finished product that was envisioned in the pre-production stage.

Editing is the main activity during the post-production. For short projects that are a couple of minutes in length, post-production editing may be completed in a matter of hours. For longer projects such as documentaries and dramas, editing may extend over a period of weeks. And for projects as TV movies, editing can take several months (COMPESI; GOMEZ, 2005).

Video editing consists of editing video footage (crop frames, order clips), adding titles, credits, transitions and captions, editing the soundtrack, adding sound effects and music, changing colour and exposure, and any other activity to improve the video product.

Video editing is supported by editing software. Some of them are normally already installed on the operating system, like MovieMaker for Windows or iMovie for Macs. Other ones are online editing software as YouTube's video editor. Furthermore, there are software products that have advanced features as XMedia Recode¹, Video To Video^{2, 3}, Premiere Pro⁴ or Final Cut⁵.

Besides the editing activity, Kirk *et al.* (2007) defined other activities as leaving content on tape (for some people it was part of an explicit archival practice, for others it was because they had no means by which they could edit the footage or they want to share immediately), download to the computer (as first step to start the edition), file for later retrieval (filed in folders for later retrieval), and digital archiving and backing-up of the edited films.

¹ <http://www.xmedia-recode.de/download.html>

² <http://www.videotovideo.org/>

³ <http://www.videostudiopro.com/en/>

⁴ <https://www.adobe.com/br/products/premiere.html>

⁵ <http://www.apple.com/br/final-cut-pro/>

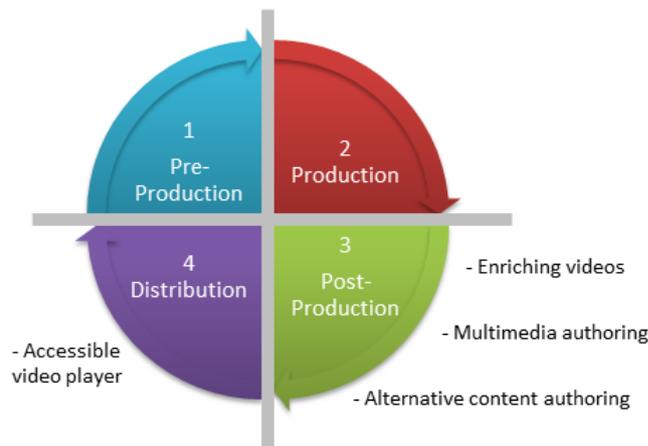
Distribution

The final stage of any project is the distribution. It consists of delivering the content to the audience. According to [Compesi and Gomez \(2005\)](#), the video may be designed and produced for distribution via website, local Intranet or at college or university. For projects that are completed, as a course assignment in a college or university production class, distribution may be as simple as watching the final project on the last day of class. For a corporate training video, distribution may mean that the program is made available on a video server to be accessed by employees via their desktop computers, connected to the central server via a local Intranet. The production considerations for video that is meant to be distributed to a worldwide audience via the web are somewhat directly on a video monitor. As different as these examples are, the ultimate goal of any production is to reach a specified audience and to effect them in some way.

2.3 Literature Review

In this section, the literature review is focused on the post-production and distribution stages, considering the video production process. [Figure 2](#) shows the four stages of the video production process, which the post-production stage highlights enriching videos, multimedia authoring and alternative content authoring, and the distribution stage highlights the accessible video player.

Figure 2 – Literature review focused on the post-production and distribution stages



Source: Elaborated by the author.

At post-production stage, activities such as downloading, video editing, and digital archiving are well defined. At this stage, research studies associated to multimedia authoring, enriching videos, and alternative content authoring need to be considered to understand this thesis.

Multimedia authoring. Multimedia authoring involves collating, structuring and presenting information in the form of a digital multimedia, which can incorporate text, audio, and

still and moving images (SHARDA, 2008). Research studies in the area are concentrated on creating models, templates, multimedia presentations, among others. Researchers usually explain different authoring paradigms and also discuss implementation languages. Hardman, Rossum and Bulterman (1993) developed an environment for constructing and playing multimedia presentation using a structured paradigm. Bulterman and Hardman (2005) also explained the construction of multimedia presentations. However, they used four paradigms: structured, timeline, graph and scripting. Guimarães, Cesar and Bulterman (2010) developed a multimedia document model which allows end-users to create and share comments on third party videos. Meixner *et al.* (2014) developed an authoring tool for the enrichment of video content with annotations and described an iterative process for improving the usability of the authoring tool. Sadallah, Aubert and Prié (2014) explained an annotation-driven and component-based model to conceptualize hypervideos. Guimarães (2014) described a social-aware multimedia authoring workflow which highlights the social aspects with social interaction through annotations. Amorim, Santos and Muchaluat-Saade (2015) developed adaptative layouts for hypermedia documents and the use of nesting templates.

Enriching videos. Enriched videos are videos augmented with visual enrichment (captions, still images, video fragments, etc.), audio enrichment (voices, sounds) or tactile enrichment (using vibrating or Braille devices) (ENCELLE *et al.*, 2011). The goal is to provide a translation or complement for people who cannot fully understand the video content. Encelle *et al.* (2011) developed audio enrichment using speech synthesis and nonverbal audio messages for blind people. Champin *et al.* (2010) proposed a collaborative annotation for video content. Rodrigues *et al.* (2010) explained additional multimedia interactive content to television programs to assist people with cognitive disabilities. Villamizar *et al.* (2011) described a concept about adaptive personalization for people with sensory disabilities.

Alternative content authoring. This category involves studies which improve alternative content (caption, audio description, transcript or sign language), and others that explain the elements which are necessary for a video to be accessible. Hong *et al.* (2011) proposed a framework for dynamic captioning, whereby the system localizes the actor speaking at any given time, and displays the corresponding speech next to the actor to increase video accessibility for the hearing impaired people. Antunes *et al.* (2011) created a framework to support sign language recognition. Lopez-Colino and Colas (2012) also created an architecture that promotes collaboration between sign experts and developers. Araújo *et al.* (2014) allowed the insertion of sign language video tracks for people with hearing disabilities, they use avatars and the final video is reviewed by experts. Carrera *et al.* (2013) include some accessibility recommendations in each stages of audiovisual product creation, focused in academic material. PAe (2014) explained some recommendations for alternative content and described some tools. Hilera-González and Campo-Montalvo (2015) explained recommendations to create accessible digital content, including

audiovisual content. The University of Washington⁶ produced a work about video accessibility. [García-Crespo, López-Cuadrado and González-Carrasco \(2016\)](#) explain alternatives to access information on cellphones.

At distribution stage, the accessible video players have an important role to video diffusion.

Development of accessible video players. Accessible video players are video players that allow navigation using different input devices (mouse, keyboard, joysticks, etc.) and assistive technologies (screen readers), and provide functionalities that users can choose according to their needs. Some researchers developed video players with several functionalities as play/pause, stop, rewinding of seconds, forwarding of seconds, volume, caption on/off, audio description on/off, help of guidance, among others ([GONZÁLEZ *et al.*, 2011](#)), ([ABLEPLAYER, 2015](#)) and ([AFB-ACCESSIBLEPLAYER, 2015](#)). Other researchers focused on specific functionalities, [Cheng *et al.* \(2009\)](#) developed a forwarding functionality that can be adapted according to the user preferences, [Li *et al.* \(2009\)](#) developed a video player with functionalities as addition of annotations, transcript's edition, searching on the transcript or the annotations, and show slides of a presentation. [Anthony, Kim and Findlater \(2013\)](#) developed adaptable video players for people with motor impairments.

Next subsections describe in more details three of these categories, closer with this thesis focus. The first subsection describes work about multimedia authoring; the second one is about accessibility in video content and the last subsection refers to accessible video players. These categories will help the reader understand which stages are part of the multimedia authoring process, which guidelines exist for alternative content and which accessible video players were developed.

2.3.1 *Multimedia authoring*

[Kirk *et al.* \(2007\)](#) and [Guimarães \(2014\)](#) developed studies with amateur users, which is the focus of this work. They explained stages involved in the multimedia authoring process for creating videos, being that the second work more focused on social interaction.

[Kirk *et al.* \(2007\)](#) described a lifecycle model to create videos based on a field study conducted with 12 families and 7 students. The model is divided in four stages: a pre-capture, at capture, post-capture and end use stage.

Each stage entails a different set of activities.

1. **Pre-capture.** The importance of an event leads users to identify a value in video. A video is a way in which users can document or celebrate special events. Users look for specialized

⁶ University of Washington - Creating Accessible Videos in <http://www.washington.edu/accessibility/videos/>

video devices according to the importance of events. The main activity is to get a capturing device, such as digital cameras or cellphones.

2. **At capture.** The three activities are: capture, edit on device and share on/from device. The quality of capturing is related to the storage and the importance of the event being captured, so that the two types of capture are spontaneous (usually using cellphones) and ad-hoc capture (using digital cameras). Editing video on devices was rare and only one participant could add titles and fades before sharing the video. When the video was recorded with cellphones, participants reported that they would often share it immediately.
3. **Post-capture.** The activities were: leaving content on tape (as explicit archival practice), download to the computer, file for later retrieval, editing on computer (crop frames, order clips, add titles, captions and transitions) and digital archiving and back-up.
4. **End use.** Sharing the footage with others.

Some issues reported by [Kirk et al. \(2007\)](#) have been solved, such as issues with limited storage, poor video quality and sharing (by email, DVD burning or posting small size video clips to a blog).

[Guimarães \(2014\)](#) describes a social-aware multimedia authoring workflow with four stages: Capture and Processing, Access and Navigation, Creation and Production, and Content Enrichment. Social-awareness is related to social aspects around personal media. They aim to help small groups of people (such as a family, a school class or a sporting club) viewing, creating, and sharing personalized multimedia. The four stages are related directly with the user interaction inside a social context to find a video and interact with parts of interest.

1. **Capture and Processing.** Based on video capture by people with strong ties to facilitate media annotation and content selection.
2. **Access and Navigation.** Consist of improving content searching and selection by individual users of a shared media repository.
3. **Creation and Production.** Content authoring becomes an incremental process of content refinement, sharing and repurposing. They focus on personalized video stories based on the social bonds between people. They developed a system to compose medias and adding annotations.
4. **Content Enrichment.** Media can take on new meaning based on the insights of viewers. The authors propose allow any user - not necessarily the author - to add comments.

Only two research studies about multimedia authoring process were found in the literature. [Kirk et al. \(2007\)](#) described four stages of video production process based on field study with

users, however the only activity regarding accessibility cited was the addition of captions at post-production stage, but they did not described any detail about this activity. [Guimarães \(2014\)](#) explained four stages including content enrichment. As they focused on social interaction, they did not incorporate any accessibility issue during the process. The main activities were the composition of media in the production stage, and the content enrichment through the addition of comments to interact with other users.

2.3.2 Accessibility in video content

Following are described guides which explain the aspects necessary for a video to be accessible.

[Carrera et al. \(2013\)](#) include accessibility recommendations in each stage of audiovisual product creation (planning, production and diffusion) and explain the Spain laws about recording authorizations and Creative Commons licenses. As they are focused in academic material, they present examples about recording authorizations and scripts (one script include captions and audio description). They explain some recommendations about captions and audio description, but they did not explain steps for alternative content generation, and the recommendations are not explained with details.

The *Dirección de Tecnologías de la Información y las Comunicaciones* of Spanish Government ([PAE, 2014](#)) explained accessibility recommendations for web developers to include audio and video in a web site. They explained how to include a video using HTML5 and Flash, definitions and examples of transcripts, audio descriptions and captions, and general accessibility criterion for multimedia content. The authors explained a general vision about what elements are necessary to make an accessible video, but did not explain how to create the alternative content.

[Hilera-González and Campo-Montalvo \(2015\)](#) explained recommendations to create accessible digital content (documents on Word and PDF, presentations on PowerPoint, audiovisual materials, and web pages), focused on academic material. They explained that an accessible video should have an alternative content (transcript, caption, audio description, and sign languages) and some functionalities of a video player to support people who need them. They describe definitions of transcript, captions (with recommendations to create captions based on the norm UNE 153020:2005 of Spain), audio description (also provide recommendations based on the norm UNE 153020:2005), and sign language (provide recommendations based on the norm UNE 139804:2007 of Spain). In addition, they explained tools, which ease the creation of audiovisual materials. Finally, they presented a list to test video accessibility for teachers' material. The authors presented recommendations based on the norms of Spain, but it is still missing transcript recommendations and as they are focused on academic material, they did not explain the video production process.

The University of Washington⁷ has an extensive work about video accessibility. They explained how to create alternative content, developed an accessible video player (explained in the next section), and developed a software⁸ that enables deaf and hard-of-hearing Americans to use sign language over a cellphone.

Most of the researchers explained accessibility recommendations for audiovisual content focused on academic material. However, they did not explained with details, that means, explaining which activities are necessary to create alternative content, which guidelines should be followed, and which tools provide support for the activities.

2.3.3 Accessible video players

The purpose of an accessible video player is to allow users with disabilities to access video content through assistive technologies. Therefore, companies and researchers have developed accessible video players. There are desktop and web-based video players. Examples of desktop video players are VLC⁹ (used to stand for VideoLAN Client), BS.Player¹⁰, Media Player Classic¹¹, among others. Most of the web-based accessible video players have used HTML5, because it provides basic controls and enables the addition of other controls. Different controls are provided, depending on the web browsers. For example, Firefox's controls provide the following functionalities: play/pause toggle, timeline with direct jump to time offset (seeking), timeline display playback position and buffer progress, playback time display, volume slider and volume on/off button (PFEIFFER, 2010).

YouTube is the most popular video player on the Web. It uses HTML5 player by default and provides controls of play/pause, volume, captions, caption options (font color, font size, font family, background color, among others), screen modes, automatic transcript, speed, quality and keyboard navigation. The social network Facebook provides controls of play/pause, full screen, volume control, caption, keyboard navigation, and option to upload captions in different languages with .srt format.

Lourdes Moreno developed an accessible HTML5 Media Player¹² using HTML5, JavaScript and CSS. The player displays the following controls: play/pause, stop, rewinding of seconds, forwarding of seconds, volume, audio, caption on/off, audio description on/off, help of guidance, selection of caption language and support to keyboard navigation. Mike Wald developed Synote player, which includes the following controls: play/pause, stop, preview and next video and volume. It also enables the addition of annotations to the synchronized multimedia (notes, tags, bookmarks, images/slides, links, and captions) (LI *et al.*, 2009).

⁷ University of Washington - Creating Accessible Videos in <<http://www.washington.edu/accessibility/videos/>>

⁸ ASL - <<http://mobileasl.cs.washington.edu/index.html>>

⁹ <http://www.videolan.org/>

¹⁰ <http://www.bsplayer.com/>

¹¹ <https://mpc-hc.org/>

¹² Lourdes Moreno's Player is available on <<http://labda.inf.uc3m.es/LourdesPlayer/>>

Terrill Thompson developed an accessible video player, at Washington University, called Universal Media Player (UMP). However, he changed the name to Able Player¹³. Able Player was developed with HTML5, JavaScript, CSS and PHP and it provides the following controls: play/pause, stop, rewind/forward, volume, closed captions on/off, audio description (creating two versions of the video, one with audio descriptions and one without), transcript, sign language, speed, chapters, player preferences to save the user's preference (for example, set closed caption on by default), help, full screen, search (using the data-search, it does not has a specific control) and support to keyboard navigation.

Floe Project¹⁴ developed a player with play/pause, closed captions on/off, transcript on/off, full screen and keyboard navigation controls. The video player of Standford University was developed with HTML5 and it offers the following controls: play/pause, speed, closed captions, full screen, previous and next video, information (keyboard controls), settings (for changes in the video player) and keyboard navigation.

VideoJS¹⁵ presents the following controls: play/pause, closed captions, volume and full screen. It also enables changes in the color of the design player. One issue using VideoJS player is that the control bar disappears after the video is played, thus using the keyboard navigation is impossible to achieve that control bar.

Other video player recognised by disabled users is the BBC's iPlayer¹⁶. The BBC iPlayer only works in the United Kingdom and it has been built with accessibility, offers clearer visual design for buttons, good contrast, keyboard and mouse navigation, support screen reader users, and support the addition of alternative content as subtitles, sign language and audio description. The AFB accessible player¹⁷ was developed with HTML5 and CSS and has the following controls: play/pause, stop, rewind/forward, volume, full screen, shortcuts and keyboard navigation. It also enables the configuration of the size and color scheme and it can be used as an embedded YouTube player. Finally, MediaElement¹⁸ was developed with HTML5 and CSS and it presents the following controls: play/pause, volume, closed caption on/off and full screen.

These accessible video players were developed in the last years, providing functionalities that help people with disabilities. Users should choose which video player is more suited to their needs. [Table 2](#) shows the video players functionalities.

¹³ Able Player is available for download on <http://ableplayer.github.io/ableplayer/>

¹⁴ Floe Project is available for download on <http://www.floeproject.org/>

¹⁵ VideoJS is available for download on <http://www.videojs.com/>

¹⁶ The BBC's iPlayer is available on <http://www.bbc.co.uk/iplayer>

¹⁷ The American Foundation for the Blind's Accessible HTML5 Video Player (AFB Accessible Player) is available for download on <http://www.afb.org/info/programs-and-services/technology-evaluation/creating-accessible-websites/download-afbs-accessible-html5-video-player/1235>

¹⁸ MediaElement is available on <http://mediaelementjs.com/>

Table 2 – Video players functionalities

Controls	You-Tube	Face-book	Mo-reno	Sy-note	Able	Floe-project	Stand-ford	Video JS	AFB	Media Element
Play/pause	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Stop	-	-	Yes	Yes	Yes	-	-	-	Yes	-
Rewind/forward	-	-	Yes	-	Yes	-	-	-	Yes	-
Volume controller	Yes	Yes	Yes	Yes	Yes	-	-	Yes	Yes	Yes
Full screen	Yes	Yes	-	-	Yes	Yes	Yes	Yes	Yes	Yes
Caption	Yes	Yes	Yes	-	Yes	Yes	Yes	Yes	-	Yes
Caption options	Yes	-	-	-	-	-	-	-	-	-
Audio description	-	-	Yes	-	Yes	-	-	-	-	-
Help	-	-	Yes	-	Yes	-	Yes	-	Yes	-
Caption language	Yes	Yes	Yes	-	-	-	-	-	-	-
Transcript	Yes	-	-	-	Yes	Yes	-	-	-	-
Sign language	-	-	-	-	Yes	-	-	-	-	-
Speed controller	Yes	-	-	-	Yes	-	Yes	-	-	-
Chapter/Annotation	-	-	-	Yes	Yes	-	-	-	-	-
Search	-	-	-	-	Yes	-	-	-	-	-
Keyboard navigation	Yes	Yes	Yes	Yes	Yes	Yes	Yes	-	Yes	-
Preference	-	-	-	-	Yes	-	-	-	-	-

2.4 Discussion

In this chapter, concepts about accessibility and video production process have been presented. These concepts are necessary to understand [chapter 3](#). Also, the chapter presented work related to this thesis, that are, work about multimedia authoring to understand where the research is focused on, work about video accessibility to know the goal of the actual researchers, and work about accessible video players, because it is the main tool for the final stage of video production process.

The video production process presented on [section 2.2](#) involves details about how to prepare, capture, edit a video, and also includes concepts about camera directions. This process is focused on professional authors who planned the capture. However, they did not consider accessibility issues that can be avoided taking into account in early stages. For example, this process did not explain how to develop alternative content for videos, neither how to prepare

the script to support alternative content. On the other hand, researchers of multimedia authoring process described activities which did not incorporate accessibility. They focus was formalize a process based on field study with users (KIRK *et al.*, 2007) and describe a process based on social interaction (GUIMARÃES, 2014). Therefore, a video production process that considered accessibility as part of their activities would help people with disabilities to understand the video content. The next chapter explains how to include accessibility into the video production process.

AN ACCESSIBLE VIDEO PRODUCTION PROCESS

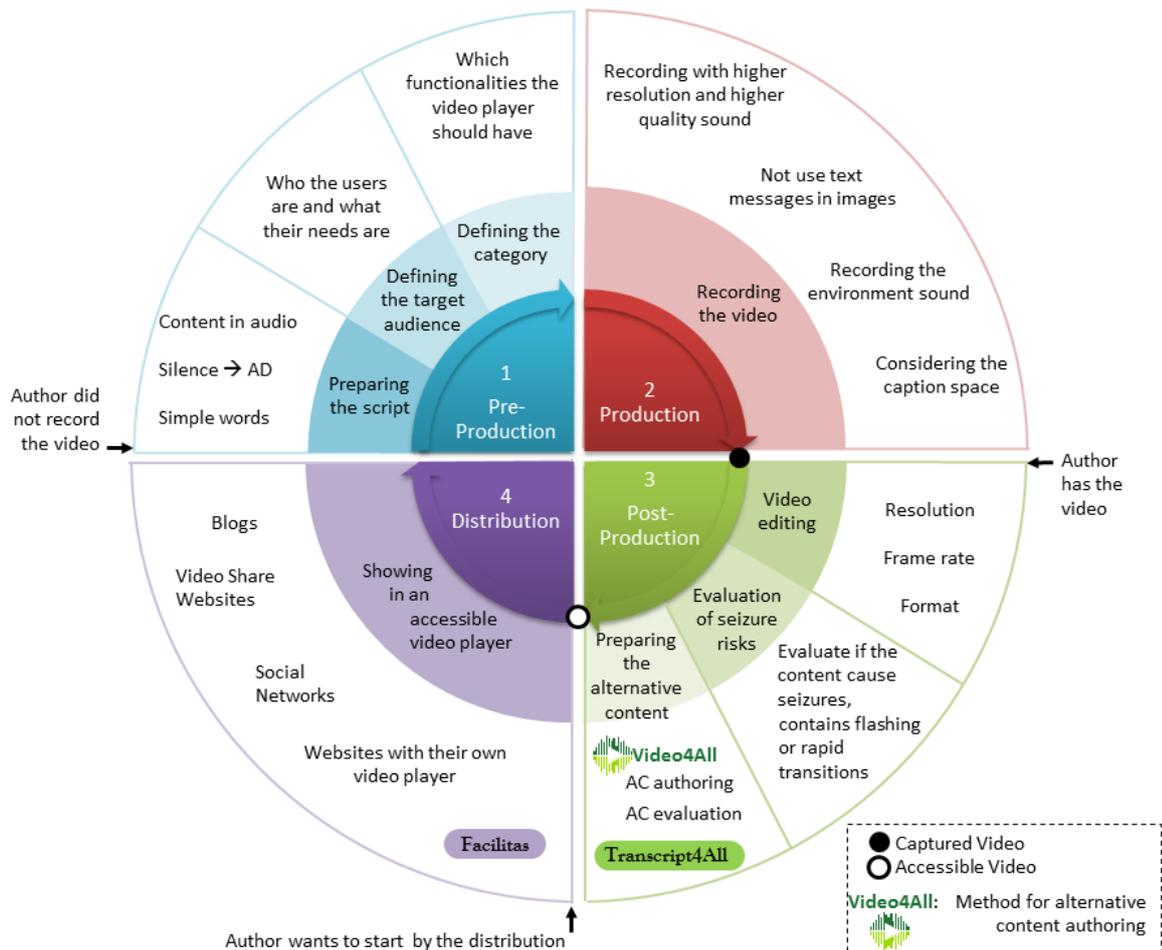
The previous chapter described accessibility guidelines to be applied on videos, research studies about video accessibility and the video production process. However, the video production process did not incorporate accessibility concerns in its stages. As [Dias \(2014\)](#) expounded, the necessity to include accessibility in the web development process is important to support web developers with techniques to produce a web system with accessibility. Similarly, [Carrera *et al.* \(2013\)](#) explained that the incomplete access to the video information is a barrier for many people (with or without disabilities). For these reasons, accessibility should be included from the beginning of the video production process.

In this chapter, the rationale for characterizing the process and the video production process are explained. The [section 3.1](#) presents the rationale for characterizing the process based on preliminary studies with users. The video production process, explained in [section 2.2](#), was adapted in order to incorporate accessibility in all stages of the process. Thus, the new process is named accessible video production process. This process is composed by four stages: pre-production ([section 3.2](#)), production ([section 3.3](#)), post-production ([section 3.4](#)) and distribution ([section 3.5](#)). Each stage has a set of activities, accessibility recommendations and tools to help the implementation of these recommendations. Due to the facility to generate video content by users on the Internet, not only professionals develop video products, but also people who may not have knowledge about it. These people can be called nonprofessionals or amateurs. This process helps amateur authors starting a new video project considering accessibility.

[Figure 3](#) shows the four stages that compose this process and their activities. Pre-production is composed by three activities: preparing the script, defining the target audience and defining the category. Production has only the role of recording the video. Post-production is responsible for three activities: video editing, validation of seizure risks and the Video4All method. The Video4All is a method for alternative content authoring, it is composed by activities

related to the alternative content authoring and evaluation. The Video4All was developed to help amateur authors. Finally, the distribution stage is composed by the activity of presenting in an accessible video player. If the author did not record the video, he/she can start by the pre-production stage. If the author has the video (recorded by himself/herself or by other authors), he/she can start by post-production stage. Moreover, if the author has the video and wants to show or share it, he/she can start by the distribution stage.

Figure 3 – An accessible video production process



Source: Elaborated by the author.

3.1 Rationale for characterizing the process

Preliminary studies with real users were conducted to collect their necessities and difficulties related to video accessibility. In this research, it was developed the Facilitas, an accessible video player (see subsection 3.5.1).

To develop Facilitas, the User-Centered Design (UCD) process was used. Four versions (see Table 3) were created increasing functionalities and solving problems detected in the tests. The UCD process comprises three phases, namely design research, design, and design evaluation.

During research, the designer aims at assessing the users and their needs. In the second phase, according to the findings from the design research, user interface, document, and other artifacts are developed. Once a design has been drafted, the UCD practitioner evaluates it with users and revises it based on the results of the evaluation (WILLIAMS, 2009).

Following are explained the four iterations of UCD:

1. Iteration 1: tests with 8 students and two (2) older people, an initial prototype was developed (version 0.1).
2. Iteration 2: tests with 15 users (adults and older people), the second prototype was developed (version 0.2).
3. Iteration 3: tests with two (2) users with auditory disability, the third prototype was developed (version 1.0).
4. Iteration 4: tests with one (1) blind user, the fourth prototype was developed (version 1.1).

Table 3 – Facilitas Player Versions

	version 0.1	version 0.2	version 1.0	version 1.1
play/pause	yes	yes	yes	yes
speed control	yes	yes	yes	yes
search	yes	yes	yes	yes
closed caption	yes	yes	yes	yes
color caption	yes	yes	yes	yes
font-size caption	yes	yes	yes	yes
font-Family caption	yes	yes	yes	yes
volume control	yes	yes	yes	yes
full screen	yes	yes	yes	yes
annotations	yes	yes	yes	yes
light	yes	yes	yes	yes
stop	no	yes	yes	yes
controls names	no	yes	yes	yes
help	no	yes	yes	yes
background color	no	yes	yes	yes
Added “go to”	no	yes	yes	yes
forward/rewind	no	no	yes	yes
DOM navigation	no	no	yes	yes
audio description	no	no	no	yes
sign language script	no	no	no	yes

3.1.1 Iteration 1 - students and older people

The first iteration reports the preliminary users' inquiry to collect requirements regarding their needs. The first prototype of Facilitas Player was developed. This prototype was tested with 10 users. This section is based on the following paper:

Villena, J. M. R.; Ramos, B. C.; Fortes, R. P. M.; Goularte, R. Web videos – concerns about accessibility based on user centered design. Procedia Computer Science , v. 27, n. 0, p. 481 – 490, 2014. ISSN 1877-0509. 5th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-exclusion, (DSAI) 2013.

Phase 1 - Design Research

The purpose in this phase is to assess users and their needs. It was performed a preliminary users' inquiry on how they watch videos on the Internet and which functionalities they use. Nine volunteers (5 males and 4 females) between 23 and 63 years, participated of the inquiry. Two participants only watched TV and 7 were computer-savvy users with experience in watching videos on the Internet. The participants were asked about what types of video programs they would like to watch. The first five types were TV shows, films, video clips, documentaries, and videos about cooking. Also, they were asked about the video players they usually used. The first five players were YouTube, Windows Media Player, Media player classic, VLC and a player from a news web site. Most of the participants used basic controls of media player, namely play/pause, resize, volume, caption and caption settings.

They were asked about problems they usually faced when they wanted to watch a video. Two participants reported problems with codec, two (older people with auditory impairment) explained that with low volume is common when they watch movies, and two with rewinding some part to hear it again. Other problems were difficulty in associating subtitle with video, errors with special characters in subtitles, caption speed, difficulty in understanding subtitles, difficulty with keyboard navigation, lack of sound equalization, lack of shortcut keys for full screen, lack of close button and automatic loading of the video.

They were asked about possible improvements in the video player. They answered they would like to edit the captions, access the keyboard, search a scene, select an interval, share some parts of the video, use voice recognition for controls, use annotation on the video because they usually needed to write the timestamp, read some description of the video to know the content and search by content rather than by title.

Phase 2 - Design

This phase involves brainstorming, conceptualizing and sketching initial drafts of the design based on findings from the design research.

The object of study was Facilitas player, an accessible video player. Although the main content is the video, alternative content, such as closed caption, transcription, audio description, and sign language complements it. For a first iteration, it was used video and caption as content and the users' data collected in phase 1. It was analyzed the guidelines proposed by UAAG 2.0, WCAG 2.0 and ISO 9241-171 to develop the functionalities for the first version of the prototype. Such functionalities are play/pause, speed controller (slower and faster), search of text on caption, closed caption on/off, settings (toolbar position, and change in font family, font color and font size), volume controller, full screen, tags (link to a specific time in the video), light controller, keyboard navigation, emphasis on the selection, and focus on the use of the keyboard. The Facilitas player controls are shown in Figure 4.



Figure 4 – Facilitas Text Configuration

It was defined tag as a link to some specific time in the video. In phase 1, a user reported she wanted to use annotation on the video because she usually writes the timestamp. The tags are composed of three elements, namely name of the tag (for example, in Figure 5, “Tip: dark chocolate”), description, when the tag is selected (in Figure 5, “01:00 - how to make dark chocolate”) and marks in the timeline (Figure 5, tag “Tip: dark chocolate” is the third mark in the timeline). The tag control enables the developer to add tags to the video, which facilitates the search. For instance, in Figure 5, the video has six tags. If it is moved the mouse over the “Tip: dark chocolate” tag, a long description appears; if the tag is selected, the video skips to minute 01:00.

The search control enables the search for a word or phrase that appears in the subtitle. The player will show all results in a list and when one is selected, the player reproduces the selected part of the video. For instance, in Figure 5, it is searched for the word “popular” and one result was returned. When this result is chosen, the player reproduces the video.

Finally, the light controller, represented as a lamp icon on the video, enables the player



Figure 5 – Facilitas Search Control and Intentional Annotation Control

to turn on/off lights. When the light is on, the player is positioned in the middle of the screen and hides the rest of the page content, emphasizing the player, see Figure 5. This feature can help people with visual impairment as the brightness outside the player decreases.

Phase 3 - Design Evaluation

Five videos with Portuguese subtitles were used for the testing (two with English audio and three with Portuguese audio). Between 2 and 7 tags (5 on average) were created for each video. The videos lasted 4 to 10 minutes (6 on average), see Table 4.

Table 4 – Features of videos

Video type	Audio	Subtitle	#Tags	Time
Cooking	Pt	Pt	7	4:25
Documentary	Pt	Pt	6	10:17
Comedy TV show	En	Pt	4	5:58
Terror	En	Pt	2	4:29
Comedy	Pt	Pt	5	4:08

Two applications of Morae software (3.3.3 version) were used, i.e. Recorder and Manager, to facilitate the research process and data analysis. This version was used in all iterations. Tests were performed in an HP pavilion dv5 with Windows 7 (AMD Turim II P520 Dual-Core Processor 2.30 GHz) and Google Chrome browser.

An experiment to know the functionalities participants chose to complete the tasks, was conducted. All participants were provided with five videos and instructed to choose two of them.

They also answered three questions for each video. The first two tasks were questions about video content. For each question, they had to show the scene where the answer could be found. They used some search, tags, speed controller and timeline to find the scene. In the third task, they used the settings panel to configure color, size and font.

Ten individuals (six university students) whose age ranged between 23 and 63 years participated in the experiment (See Table 5). Four participants were female. Each participant spent approximately 30 minutes on the experiment. They chose two videos based on their preferences: 35% chose a cooking video, 30% chose a documentary, 20% chose a comedy TV show, 10% chose a terror movie, and 5% chose a comedy movie. Although participant 7 has a mild hearing impairment, she did not use any hearing aid. Three users (P1, P2 and P4) have myopia and wore glasses, which did not affect the interaction. The other participants have no disability.

Table 5 – Participant’s Characteristics

Participant	Age	Gender	First Video	Second video
P1	26	M	TV show	Comedy
P2	27	F	Cooking	Documentary
P3	23	F	TV show	Cooking
P4	24	M	Documentary	Cooking
P5	32	M	Documentary	Cooking
P6	25	M	TV show	Terror
P7	59	F	Terror	Documentary
P8	63	M	Documentary	Cooking
P9	23	M	Cooking	Documentary
P10	23	F	Cooking	TV show

The time it took each participant to complete a task, was analyzed. The tasks were labeled with a letter (P=participant, V=video, T=task) followed by a number. Tasks P1V1T2 and P2V1T1 required more time than the other tasks. In P1V1T2, P1 searched eight times with words that did not return the expected result - he used the timeline to complete the task. P2 also said she did not pay attention to the video because she was distracted using the video player controls. In P2V1T1 task, she used four tags and one search, but she did not perform the task; then she used the timeline to complete the task. After the answer had been found, she tested search control using one word from the subtitle and checked if the search result was correct.

P1 spent less time on the second video than on the first. He paid attention and answered the question about the video content (P1V2T1) using his memory. He used one tag in P1V2T2. Some tasks required more time than others, independently of the participant’s characteristics. For example, in the terror movie, task 2 required more time than task 1. Both P6V2T1 and P7V1T1 used the timeline twice to complete the task. P6V2T2 and P7V1T2 used search three times to complete the task. In the P7V1T2 task, the participant wasted her time on misspelled words.

Two women (P3 and P10) chose a comedy TV show. It took both participants longer to complete the first task. In the P3V1T1 task, the participant used two tags and one search to complete it. In P10V2T1, she used the same word of P3 in the search. In both P3V1T2 and P10V2T2 tasks, the participant used one search to complete them.

70% of the participants (P1, P2, P3, P5, P7, P8 and P10) spent more time on tasks for the first video than on those for the second video, i.e. they learned how to use the control functionalities in the first video.

To know the controls used by the participants during a task, each video was analyzed and counted how many times they used each control until the task had been completed. The search control was used 53,6%, timeline 30,9%, tags 13,4% and speed controller 2.1%. Also it was counted how many times the participants used each control to complete the task. The search control was used 62,5%, timeline 30%, speed controller 5% and tags 2,5%.

The search control was the most useful control, followed by the timeline. Although participants used tags, only one task was completed with them. Participants also used speed controller to complete two tasks. After having watched V1, P1 said "one thing that was different was the marking of tags. I had never seen this functionality before, but I paid attention to see what happened". After watching V2, P1 wanted to add new tags to the video and said "Adding tags to the video would be a good option. I would like to rename the first tag". P3 said at the beginning she did not understand the tags, but then she realized they were marked in the timeline. P9 said tags mean keywords and made a relation with the keywords used in YouTube. All participants, except P1 and P3, suggested they could have had the option of adding some tags to the videos. P4 and P5 suggested deleting some tags.

Regarding the caption configuration task, the participants were asked to perform one task for video 1 and another for video 2. Three types of tasks were created: change font style to Verdana, change font color to yellow, and change font size to 20. Every change in the caption configuration is automatically showed in real time and with no button. A feature of the panel settings is the automatic change in style, color and size, with no click on the button.

Five Participants (P2, P5, P7, P8 and P10) had difficulties in changing the subtitle color. By default, the point (see in Fig. 1) was in the upper left corner (white color), therefore, when the participant changed the color, the subtitle was always white, which caused some confusion. P2, P5 and P10 have computing skills, but P7 and P8 do not. When P7 learned how to change color, she began to test changing subtitle colors. P5 liked the color change feature and P10 said the change size feature is really important.

Lessons Learned

The first iteration was tested with 10 real non-disabled users. Some problems reported by users can be solved in the pre-production stage. Caption speed can be considered in the

recommendation “analysis of captions speed”. Difficulty in understanding subtitles, must be considered in the recommendation “simplicity of words”. The low volume is common when participants watch movies, the recording with high quality sound must be considered when the capture is planned.

The experiment provided important results about controls, tags and search, which will help to improve Facilitas Player. The tag control was not frequently used in the test cases, probably because it was different from other functionalities and the participants had never seen it before. On the other hand, search control seems to be very useful, as it was used in 62,5% of the cases for a successful completion of a task. Search functionality was more intuitive than tags.

During the test, a misunderstanding regarding tags and keywords were detected, therefore, a changing the name 'tag' to 'link', and the addition of the phrase 'Go to:' before the tags to guide the participants were proposed. These changes were applied in the second iteration.

3.1.2 Iteration 2 - adults and older people

The second iteration was tested with adults and older people, so that some functionalities of video player common for them were found. This section is based on the following paper:

Villena, J. M. R.; Ramos, B. C.; Fortes, R. P.; Goularte, R. An accessible video player for older people: Issues from a user test. Procedia Computer Science , v. 27, n. 0, p. 168 – 175, 2014. ISSN 1877-0509. 5th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-exclusion, (DSAI) 2013.

Phase 1 - Design Research

The purpose in this phase was to assess users and their needs, therefore, a preliminary user survey with 15 volunteers was performed. This survey is about how older users watch videos on the Internet and which functionalities they use. The participants were monitored for two months at a local Educational Institute for the elderly that offers them courses of Informatics.

They were invited to answer a questionnaire with 18 questions among which one regarded video players more commonly used. The first five were YouTube, Globo News, Terra News, UOL News and Google. Most participants used basic controls of the video players, namely play/pause, stop, change the caption size and full screen. Another question regarded the problems they commonly face when they watch a video. Eight participants reported problems with language (53%), five explained that the caption size is small (33%), five reported that the video had low volume (33%), five said that the caption was unavailable (33%), four reported the low video quality when watch the video as full screen (26%), two complained about the fast caption (13%) and two reported problems with plugins (13%). Other problems reported were contrast of captions, small screen, rewinding to some part of video to hear it again and fear to make mistakes and close everything.

Phase 2 - Design

Some problems were detected in the second prototype and in this phase they were solved. First, the name 'tag' was changed to 'link'. The links maintained the three elements: name of the link, description when the link is selected, and marks in the timeline. Second, the phrase "Go to:" was added before the links and told the participants they were to be clicked, see Figure 6.

Two functionalities were added to the setting panel. First, the player enables the setting of the background color of the subtitle. And second, the player currently has two languages, Portuguese and English, see Figure 6.



Figure 6 – Facilitas Caption Background Color and Link Control

Two accessibility features were added. First, the keyboard shortcuts, located at the Help Control, see Figure 7, then the control description to facilitate navigation (Figure 8).

Phase 3 - Design Evaluation

An experiment to know the functionalities chosen by the participants to complete tasks was conducted. Participants were provided with the same five videos used in the first iteration (see Table 4). Morae Software and the same notebook used in the first iteration to capture the audio, video, user input, and on-screen activities, were used.

Fifteen older people, whose age ranged from 40 to 82 years, participated in the experiment (Table 6). Ten participants were female. Each participant spent approximately 50 minutes on the experiment. They chose two videos based on their preferences. Most of them have visual impairments, as shown in Table 6.

Participants P1, P4, P9, P13 and P14 had used computers for over 6 years, P5, P7 and P12 between 2 and 4 years, P15 between 4 and 6 years, and the others for less than 2 years.

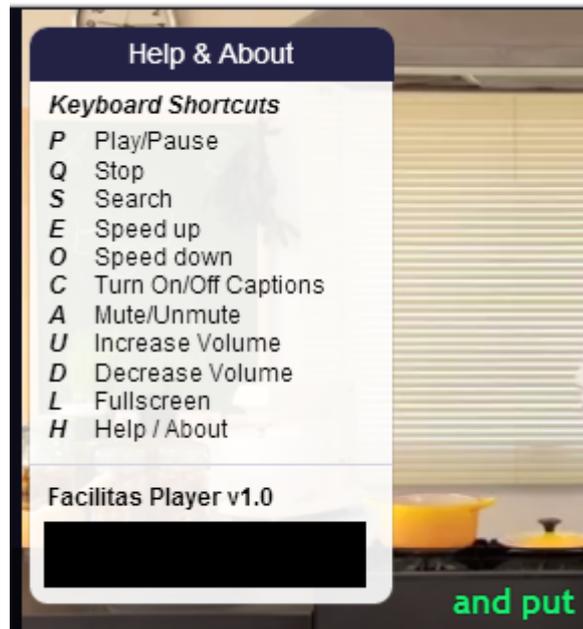


Figure 7 – Facilitas Help Control



Figure 8 – Facilitas Controls Names

Table 6 – Participant's Characteristics

Participant	Age	Gen.	Visual	Hearing	Motor	Cognitive
P1	63	M	Some	-	-	-
P2	82	F	Some	Some	Some	Some
P3	66	F	Some	-	-	Some
P4	70	M	Some	-	-	-
P5	62	F	Big	Big	-	Some
P6	59	F	Some	-	-	Some
P7	60	M	Some	-	-	-
P8	80	F	Big	Some	-	Some
P9	66	M	Some	-	-	Some
P10	66	F	Some	Some	-	-
P11	45	F	Some	Some	-	Some
P12	40	F	-	-	-	-
P13	67	M	Some	-	-	Some
P14	59	F	Some	Some	-	-
P15	52	F	Some	-	-	-

Below are the main issues related to their personal difficulties during the test:

- Regarding cognitive disability, their main difficulty was memorization. P3 has difficulty in memorizing acronyms (msn, like link in facebook, FW, CC, CCO, etc.). She reported understanding acronyms is very difficult, because the new generation's logic must be understood. P2 moves the mouse around the video player controls and can see the button's names; she loves it because she frequently forgets the control functionality. Whenever she performs a task, she reads the control name. P3 also reads controls' names and likes it. P9 said he likes to read the links' names.
- Regarding handling of computers: P6 had started to use a computer 6 months before and reported difficulties in using the video player during the test, although she reproduced the video and configured the caption text successfully. P10 said age is a problem for the manipulation of computers because the programs have plenty of information and the applications are frequently updated. P11 said she wants to learn how to use basic things, such as Skype and Google; she also said videos are important, but she needs more practice. She liked the search control of the video player.
- Regarding motor difficulties: P9 and P10 have difficulties moving the mouse.

Below are the main issues related to the use of Facilitas. All participants watched the videos with captions. The captions configuration seems to be an essential resource to them because, for example, P1 said captions are important for those who have hearing impairment. P14 liked the black border of the caption text and suggested captions could appear slowly. Controls CC (Closed Caption) and configuration were confusing to some participants (P1, P2, P4, P5, P6 and P8), because when they wanted to change the caption color or size, they did not know which control to choose. The Facilitas prototype should be redesigned regarding this issue. All participants tested the caption color background; P3, P6, P7, P12 liked black and transparent black. P13 said the caption background depends on the video background.

P1 did not use links, but he would want to add them. P3 used them and said links are the localization of a video moment. P5 said she did not add links because she needed to familiarize with the video. P13 said links are useful when he wants to return to a specific part of the video.

P1 and P12 had difficulties to identify the control to be used for searches and confused search, caption and configuration. P1 and P3 watch videos to learn languages, therefore, P3 focused on the search control because she thinks it is a tool for reviewing the meaning of a word she did not understand during her study. P15 believes the search control is faster than other controls when a person searches a specific part of the video.

P3 found useful the speed control useful for study of languages, especially because the pronunciation of words can be heard. P8 enjoys watching ballet and could use the speed down

control to see the movements. P13 enjoys the speed down control because he watches horse races and would use it.

During the test, the participants could suggest new resources to be implemented in Facilitas. P3, P9, P10, P12, P13 and P15 suggested using a new control to rewind 10 seconds and forward 10 seconds. The new controls would help the elderly to more precisely return and forward different parts of the video. P4 suggested a new control to mark some parts of the video and show them again to other people. P5 suggested a new control to capture images of video, since she likes to paint pictures. P1, P7 and P9 have problems to watch videos in languages other than Portuguese. They suggest adding a control to change the audio language. P5 has difficulty to understand how to use the video player controls and suggests making a documentation explaining each of them.

Lessons Learned

The second iteration described a research with adults and older people. Our goal was to gather impressions of older people, regarding the design of a prototype developed as an accessible video player. A user test was conducted with 15 older people and a set of issues regarding the main difficulties they faced to use the prototype was reported by them. The test also provided important results about new controls, links and search, which will help us to improve Facilitas Player. The users have enriched the requirements for an accessible video player and the current functionalities were discussed and evaluated for the participants. They have cognitive difficulties, especially regarding memory. Some controls of a video player should support them in this issue. In case of Facilitas, the link controls help to return to a specific part of a video the participant can not remember. The names of controls help users to remember functionalities they have forgotten. Participants suggest the development of a control to rewind and forward 10 seconds of a video.

The caption speed problem is repeated in the second iteration. It is important to be solved in the pre-production stage. Another problem reported by the participants is that low resolution on videos is an issue when they want to watch as full screen.

3.1.3 Iteration 3 - users with auditory disabilities

The third iteration was tested with users with auditory disabilities. Morae Software (3.3.3 version) was used. The tests were performed in a Samsung 500P4CAD3 Notebook with Windows 8 (Intel®Core™3630QM processor, 6GB of RAM, 1TB of hard drive) and a 22-inch Samsung monitor.

In the first phase, the pre-session questionnaire is explained; in the second, the Facilitas player prototype is explained and in the third phase, the test with users is described in details.

Phase 1 - Design Research

A questionnaire with 21 questions to two users was applied. Dora (not her real name) is 26 years old, single, has not finished high school, learned to use computers 10 years before, and had used the Internet for 6 years. She works at an assembly industry and uses a computer at home. She used to watch videos once a week through YouTube, Facebook, and a local news website. She uses only play control to watch videos. She declared she did not have any visual impairment, however she wears glasses (at the moment of the test she was not wearing them). She is hard of hearing since birth and also has difficulty to talk. She wears a hearing aid, but at the moment of the test she was not wearing it because the battery was dead. She is learning sign language. Two sessions with her were conducted.

Alex (not his real name) is 22 years old, deaf and single. He has never studied at regular schools, instead, he attends the Association of Parents and Friends of Children (APAE) and does not work. He learned to use the computer and Internet 3 years before. He uses a computer at home. He used to watch videos once a month through Facebook. He used only play control to watch videos. He used to ask his sister for help. He declared he has some visual impairment, but he does not wear glasses. He wears a hearing aid, but at the moment of the test, he was not wearing it. He is learning the Brazilian Sign Language known as "Libras".

Phase 2 - Design

A new version of Facilitas Player (Figure 9) was created, and anyone can use plugin to play a video on a web page.

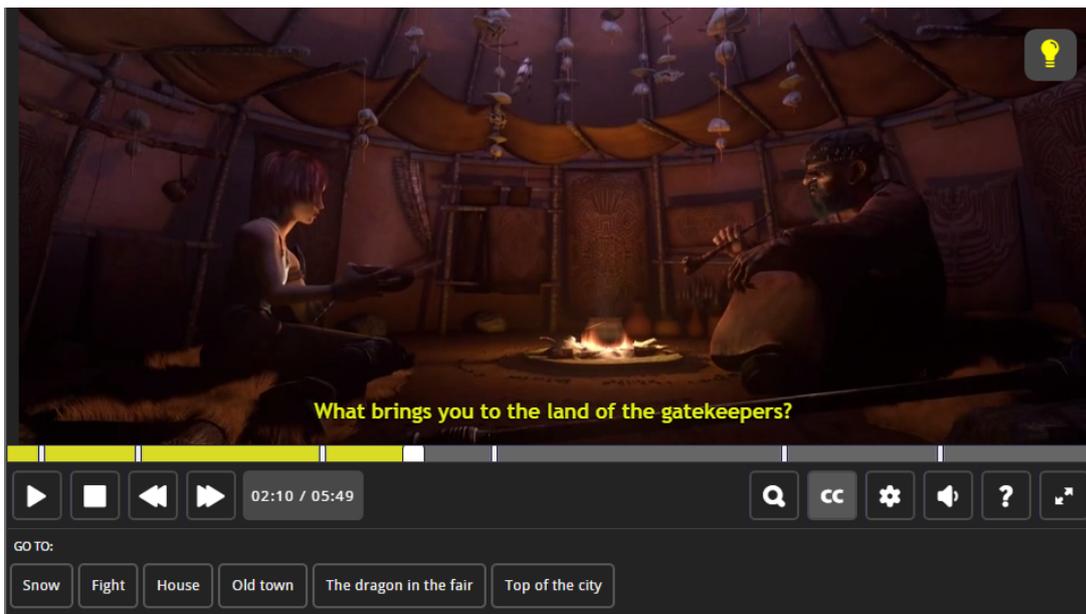


Figure 9 – New interface of Facilitas Player

Some functionalities were added: rewind 10 seconds and forward 10 seconds, as a result of the second iteration. The participants reported these features are important for the learning of

another language. Facilitas player provides structural navigation, so that the structure types are recognized by heading.

The 'links' were renamed as 'intentional annotations', which are indices participants create during an activity to mark particular time points (MINNEMAN *et al.*, 1995).

They are annotations held by the user who will use Facilitas player's plugin to play a video. Annotations facilitate the localization of a specific part of a video. Participants said they help when they want to watch tutorial videos, news or videos containing steps for the accomplishment of a task.

Another feature developed is audio description (AD), which is an audio that describes visual components of an event for users who either are blind, or have visual impairment. AD must be synchronized to the video and reproduced during the silence of the primary audio to not interrupt it.

Phase 3 - Design Evaluation

An experiment with Dora and Alex, two users with auditory disabilities, was performed. In this prototype, the audio description was not included. Two sessions with Dora and one with Alex, were conducted.

Some steps were followed in each session. First, an initial contact was made with the user; then, the researchers read and explained the procedure of the case study and the user signed the user's acceptance testing. Personal information was collected through a pre-session questionnaire. Next, the characteristics of Facilitas were explained and the user was instructed to watch two videos and answer some questions about their content. Finally, the user answered a pos-session questionnaire. Morae Software was used to capture the audio, video, user input and on-screen activities. The tests were performed in a Samsung 500P4CAD3 Notebook with Windows 8 (Intel® Core™ i7 3630QM processor, 6GB of RAM, 1TB of hard drive) and a Samsung monitor of 22 inch.

Four videos along the sessions were used. Two of them were selected (cooking and documentary) for the first two iterations because they were chosen by the participants in the first iteration. Then, two more were selected (cartoon videos Sintel and Rio), as they would be more easily understood by the users with auditory disabilities. The cartoon videos were used with Alex, because he is deaf.

A 4:25 minute cooking video showed a recipe called *brigadeiro* (VILLENA, 2014). It explains how to make *brigadeiro*, which is part of the Brazilian Culture.

The 08:55 minute documentary video was about Sao Paulo's subway system (VILLENA, 2014).

The cartoon video called Sintel showed the story of a girl and her dragon. It is an open

project of Blender (BLENDER, 2010). A transcript and closed captions for it were created (VILLENA, 2014) and used its first 5:49 minutes.

The cartoon video called Rio shows the adventures of a parrot called Blu in Brazil. The first 12:44 minutes was used.

First session with Dora

The first session with Dora lasted 1 hour and 15 minutes, see Figure 10. The pre-session questionnaire was explained in Phase 1 of this iteration.



Figure 10 – First Session with Dora

She did not like to watch the video in full screen because it affected her vision. She configured the subtitle size from 14 to 20 and changed the color to green. She wore an earphone and configured the volume; it was high, she clicked on the volume button and the volume was mute, therefore, she increased the volume 25%.

A document with questions about the video was prepared. It was used open-ended questions, as they require answers with more than one word. The document was given to Dora and she was instructed to watch the video again whenever necessary. She read the three questions, but she did not remember the answers.

First question: Which tool did the woman use to kill the man? It took Dora 3m37s to answer it. She played the video and when she saw the part that contained the answer, she paused the video and showed the answer with the finger. She did not know the name of the tool, therefore the researcher said “spear” (“lança” in Portuguese). She asked the researcher to spell the word, however she did not understand it (she said "it is difficult"). The word was searched in Google and she copied it. She wrote “spear to kill man” (“lança para mata homem” in Portuguese). She did not write it correctly, as she omitted the article before the noun and did not conjugate the verb.

Second question: What was the dragon chasing for in the city fair? It took her 5m07s to answer it. She played the video and watched it until minute 2; then she clicked on the mark of the third annotation "House" in the timeline, and returned to 1m40s. When the video was

approaching minute 2, she clicked on the mark again. She moved the mouse to the pause button, but she did not click. When the video reached 3:03, she moved the mouse over the fourth mark in the timeline “old city”, but she did not click. She continued watching the video until minute 5:06 and paused it. It was difficult to understand her answer. The question was asked by the researcher and she answered “chicken” (“galinha” in Portuguese) and wrote “galinha”.

Third question: Which part of the Dragon’s body was injured? It took her 36 seconds to answer it. She said “close to the wing” (“para perto da asa” in Portuguese) and after the researcher confirmed it, she wrote “was arm wing” (“foi braço da asa” in Portuguese). She omitted the article before the noun.

The second video (documentary about Sao Paulo’s subway system) was played. She did not wear earphones because the volume of the video was high.

First question: How many stations comprise Sao Paulo’s subway system? It took her 3m15s to answer it. She replied “4 million trains” and “4 million kilometers”. As the answer was not right, the researcher instructed her to watch some parts of the video again. She clicked on annotation “Challenges”, moved the mouse over annotation “Description”, but returned to “Challenges” and clicked on the reproduce button. She watched the video until the answer appeared, paused it and said “there are 64”. She wrote the whole subtitle “64 stations and 78 km of rails that are eerily silent”. She did not answer it objectively, but copied the subtitle.

Second question: In what city did the railroad accident occurred? It took her 6m16s to answer it. She did not remember the answer, therefore, she reproduced the video and clicked on annotation “Description”. The video was rewind some seconds and she continued watching it. When it reached the mark of annotation “Interviews” on the timeline, she used the mouse and clicked on it and the video was rewind some seconds. She continued watching it. When it reached the mark of annotation “Control Center”, she clicked on it, the video was rewind some seconds, and she continued watching it. She read the question again and continued watching the video. She clicked on annotation “rails check” and returned 40 seconds. She continued watching it until the rail road accident. She then moved the mouse to the pause button, but she did not click on it. She moved the mouse to the timeline and returned some seconds until she found the answer. She paused the video and showed the answer. She wrote the whole subtitle “in October 2000, a train that was crossing Hatfield, in the south of England”.

Third question: How many employees does the Control Center have? It took her 34 seconds to answer “a thousand, a thousand employees”. She was instructed to search on the video to confirm the answer. She moved the mouse over annotations “Description” and “Interviews”, and clicked on “Description”. She watched the video and then clicked on annotation “Control Center”. She forwarded the video some seconds and found the answer. She moved the mouse on the time bar, rewind the video some seconds, moved the mouse to annotation “Control Center”, but did not click on it. She moved it again to the pause button, waited for the answer to appear, and then paused the video. She showed the answer and said “3000 employees”. She

wrote exactly as the video subtitle “preparing for an attack, an army of 3000 employees”.

The post-session questionnaire was applied. The functionalities that Dora most liked were play/pause, rewind/forward and volume control. She said she had difficulties to use the annotations because when she clicked on them, the video jumped to some part and was replayed, and she needed to start it all again. Such difficulties were identified in the second video. Annotations were another functionality she did not like. She said she preferred to watch the video until she could find the part she was looking for and not to jump to some part of the video, no matter if the video was long. The functionality she used most was the annotations (7 times), followed by timeline (4 times). She attempted to click on the annotations 7 times. She said although she did not like the annotations, she was exploring how to use them. Therefore, another session with Dora was scheduled.

For a second session, close-ended questions will be used, whose answer is only “yes” or “no” or multiple choice. In this case, it was used multiple choice questions.

Second session with Dora

The second session with Dora lasted 40 minutes, see Figure 11. For this session, videos with less information were chose. The researchers opted for fantasy films and Brazilian culture. The video to show the functionalities of Facilitas Player was two minutes of The Age of Ice.

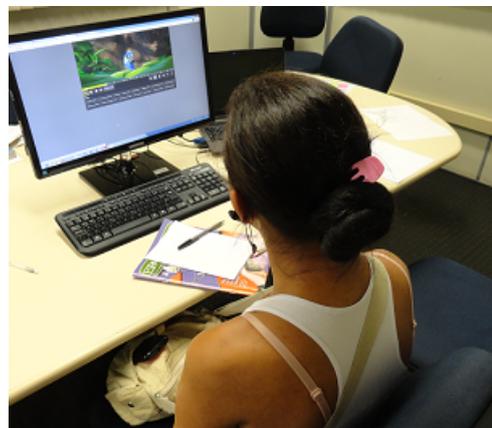


Figure 11 – Second Session with Dora

The cooking video (Brigadeiro) was played, and she listened with earphones.

First question: Does making Brigadeiro take less than 30 minutes? The answer is a yes/no type. It took her 43 seconds to answer it. She said “it is not 30 minutes, it is 15 minutes”. The researcher read part of the question “takes less than 30 minutes?”, she answered “15 minutes, not 30 minutes”. The researcher observed the question was difficult for Dora to understand, because she did not recognize the word “less”, she believed it was 30 minutes. The researcher explained the question and instructed her to choose “yes”, because she knew the answer.

Second question: What are the three ingredients used to make Brigadeiro? It took her 2m48s to answer it. She read the question, the video was paused, she clicked on annotation

“introduction”, then moved the mouse over annotation “ingredients”. She waited for 5 seconds, then the researcher instructed her to click on the play button. She clicked on annotation “ingredients” and then played the video. She watched the video and when the chef reported the ingredients, she paused it. She said “first it is condensed milk”, she wrote “condensed milk” looking at the subtitle. Then, she clicked on “ingredients” and play button. When the subtitle appeared with another ingredient, she tried to pause, but the subtitle changed. She rewound the video 10 seconds and clicked on “ingredients” again. She paused the video and wrote “butter” as a second ingredient. She played the video and when the third ingredient was reported, she paused and wrote “chocolate powder”.

Third question: When did Brigadeiro become popular? This is a multiple choice question with three alternatives: vaccination campaign, election campaign and Christmas campaign. It took Dora 1m39s to answer the third question. She said “Christmas campaign”. The researcher instructed her to watch the video again to confirm the answer. She moved the mouse over annotation “Tip - dark chocolate” and “Brigadeiro is ready”, but did not click. Dora did not remember the answer, therefore the researcher told the history of Brigadeiro. She clicked on “history” and played the video. She did not wear earphones and saw the subtitles. The researcher asked if she had understood it and if she wanted to wear the earphones. She accepted and clicked on “history” and played the video again to listen to the history. She said “election campaign” and marked the answer.

The second video (Rio) was played. She watched it, read each annotation and changed the color and size of the subtitle.

First question: Does Blu want to learn to fly? This is a yes/no question. It took Dora 10 seconds to mark “no”.

Second question: What time is the alarm set for? To facilitate the answer, the template was “__:__”. It took her 11 seconds to answer it. She clicked on annotation “alarm clock”, paused the video and she wrote the time.

Third question: How many marshmallows did Blu’s hot chocolate have? This was a multiple choice question, whose options were 4,5 and 6. It took her 41 seconds to answer it. She moved the mouse over annotation “Linda talks with Blu” and “Trip to Brazil”, but she clicked on “hot chocolate” and played the video. She moved the mouse over the pause button, waited some seconds, and clicked on pause when she found the answer. In this scene, Blu was counting the marshmallows until 5. She played the video again and Blu counted the sixth marshmallow. She paused the video and marked six.

Fourth question: What color was Linda’s coat when she found Blu in the box? This is a multiple choice question, whose options were red, blue and pink. It took Dora 1:02 minutes to answer “blue”. The researcher read the question and asked her if Linda’s coat was blue. She clicked on annotation “box in the snow” and played the video. Then she said “I think it was

about the bird”. She continued watching the video and when she saw Linda’s coat, she paused it, said “pink” and marked the correct option.

The post-session questionnaire was then applied. The functionalities Dora liked most were annotations, subtitle configuration and search. She wrote she liked annotations because they saved time. In the first session, she said she did not like the annotations functionality, but in the second session she chose it as the functionality she liked best. In the second session, she was more familiar with Facilitas Player and explored the annotation functionality. She also reported she would like other video players to have the help control functionality.

Session with Alex The session with Alex lasted 1 hour and 10 minutes, see Figure 12. He was accompanied by his cousin, who helped with the sign language. It was chose the Age of Ice film to introduce Facilitas Player to Alex and explain each functionality. When he saw the search functionality, he became excited and happy and clicked on some items of the result list. When he saw the subtitle configuration and annotations, he was surprised. He clicked on all annotations.



Figure 12 – Session with Alex

A sign language video was prepared to facilitate his understanding and added a Web Libras script¹ to the Blogger (VILLENNA, 2014).

Both videos were reproduced, i.e., the Age of Ice and a sign language video, but Alex did not want to watch the second one. He moved the head in a “more or less” signal. He preferred not to use the sign language video as his cousin said he knew how to read and write. He read and answered the questions of the pre-session questionnaire.

The first video (Sintel) was played, i. e. the same video used with Dora in the first session. The type of questions was changed and used multiple choice and yes/no questions.

First question: In which part of the body did the dragon have an injury? The options were leg, arm and head. It took him 2:45 minutes to answer it. First the question was read by the researchers, emphasising the words dragon and injury. His cousin explained the question using

¹ Web Libras - <<http://weblibras.com.br/>>

sign language, showed the parts of her body and pointed to the video. He clicked on annotations “snow”, “fight” and “home” and watched the video until he could find the answer. When he saw the dragon injured, he pointed with the pointer of the mouse. He marked the leg, instead of the arm. However, he identified the injury on the video, which means he understood the question.

Second question: What object did the woman use to kill the man in the fight? The options were knife, hammer and spear. It took him 57 seconds to answer it. His cousin explained the question using sign language and he clicked on “fight”, watched the video until the woman killed the man, paused it and marked spear.

Third question: Does the old man of the mountain house wear glasses? It was a yes/no question. It took him 1:43 minute to answer it. He moved the mouse over “snow”, but clicked on “house”, played the video and watched it. When he saw the old man, he paused the video, but the frame changed and showed the woman’s face. He clicked on the time bar to rewind the video some seconds and paused it when the old man’s face appeared. He was undecided. Therefore, the researcher instructed him to click on full screen. He clicked, paid attention and made a “no” signal with his finger.

The second video (Rio) was played, i. e. the same tested with Dora, and asked the same type of questions. When the first subtitle appeared, the researcher paused the video to configure the color and size subtitles. He changed the color twice, the size increased from 20 to 24, but he returned to 20 and watched all video. He tested the annotations.

First question: Does Blu want to learn to fly? It took him 1:24 minute to answer it. He looked at the researcher and made a positive signal with his hand, but his cousin made a “no” signal with her hand and pointed to the video. He clicked on “start” and played the video, clicked on “Blu tries to fly” and watched the scene. After one minute, he paused the video and made a “no” signal with his hand. They may have confused “want to learn” with “know” because they saw the scene when Blu tried to fly, but failed.

Second question: what time was the alarm clock set for? It took him 25 seconds to answer it. He moved the mouse over "man slips in the snow" and "Blue bird is captured", clicked on "alarm clock" and pointed to the time smiling.

Third question: How many marshmallows did Blu’s hot chocolate have? It took him 2:20 minutes to answer it. He made a “more or less” signal with his hand and the researchers instructed him to look at the video. He clicked on “Blue bird is captured” and “breakfast”, then played the video until he saw Blu talking about the marshmallows. He paused the video and played it again until Blu counted the marshmallows. He paused the video, smiled, indicated “six” with his fingers and answered the question.

Fourth question: What color was Linda’s coat when she found Blu in the box? Alex’s cousin explained the question using sign language. It took him 2:09 minutes to answer it. He searched the scene, clicked on the time bar, moved the mouse over annotation “Blue bird is

captured”, clicked on annotation “box in the snow”, and found the scene. His cousin made a “which” signal, Alex pointed to the red color, his cousin made a “no” signal, then the researcher instructed him to watch the video. He played the video and when he saw Linda’s coat, he made a “more or less” signal. He did not distinguish between red and pink. The researcher instructed him to mark any. He marked pink.

Lessons Learned

The third iteration explained the behaviour of the people with auditory disabilities when they watch a video. A user test with two users was conducted and performed three sessions. They faced some difficulties to interpret the questions. In the second session, the questions were changed and used simple present tense.

During the preparation of the videos to be tested with Alex, the researchers created sign language videos generated from an automatic translator. As the sign language changes according to the region of the country, the automatic translation may not be adequate for some users. Alex did not want to watch the sign language video. For some individuals with auditory disabilities, sign language is the first language and they may not read a written language fluently.

Providing important information in sign language and using simpler texts supplemented by images, graphs, and other illustrations help make web content more understandable to many. However, not all people with auditory disabilities know sign language (ABOU-ZAHRA, 2012).

3.1.4 Iteration 4 - blind user

A case study was conducted to understand how a person with visual disabilities might use a resource of assistive technology or other means to watch and comprehend video content, accomplish tasks and recover from task failures using workarounds. Our goal is to report the problems and situations that blind users face when watching video content on the websites. During the test, the Facilitas Player was used, but the user always compare with his experience. This section is based on the following paper:

Rosas-Villena, J.; Ramos, B.; Goularte, R.; Fortes, R. Video accessibility on the most accessed websites - a case study regarding visual disabilities. In: ANTONA, M.; STEPHANIDIS, C. (Ed.). Universal Access in Human-Computer Interaction, 17th International Conference on Human-Computer Interaction (HCII). Access to Today’s Technologies. Springer International Publishing, 2015, (Lecture Notes in Computer Science, v. 9175). p. 231–241. ISBN 978-3-319-20677-6.

Phase 1 - Design Research

An interview with David (not his real name) was conducted. David is 31 years old, married, pursuing a Bachelor in Information System degree in a undergraduate distance education

course. He is a blind software developer and he usually has access to the videos on the Web. The websites he is used to watch videos are YouTube and ESPN – Entertainment and Sports Programming Network – (he is interested in news about soccer). When playing videos, he explores all functionalities that he has access.

When David studied the subjects of Information System course, he developed a video player that supports closed caption functionality, used specially for people with auditory impairment.

He was asked about the difficulties that he usually has when accessing videos at the Internet. He said *“Some time ago I thought YouTube invested in an interesting way to provide accessibility in a video player, but sometimes I go there to see something and it is difficult to access the videos because the player is implemented in Flash, and the controls do not have any accessibility.”*

He also said *“One thing I miss in many players is the possibility of forward or rewind the video, to control the playback. For example, when you are hearing something and lose what the person spoke, you cannot listen again, then you have to click to play the video from the beginning to get to that point where the person was talking about. This is something that I miss.”*

He was asked about how accessibility works with closed captions, he said:

- *“If it is accessible, no, honestly I do not remember a player that it is accessible in the popular websites, which have an accessible caption for visual disabilities. The YouTube used to have that feature, I was able to read the captions, but after a while, I was not anymore. I do not know if YouTube changed his player code or if the screen reader I was using does not had support or does not work well with YouTube Flash applications. Currently I cannot access it anymore”*. He used the NVDA as screen reader.

Phase 2 - Design

In this phase, the audio description is the most important element to be testes, because it is a description for blind people. Also as David is a web developer, he can provide important comments about the Facilitas development.

Phase 3 - Design Evaluation

Morae software² was used to record the interview and the interaction with Facilitas Player. To test, a researcher’s notebook with Windows 8 was used and NVDA screen reader. He used to set up the configurations. The researcher’s notebook keyboard has some differences with his notebook, it took some time to recognize the key position and when he asked about an specific

² www.techsmith.com/morae.html - Morae is a software solution for usability testing, including recording of the screen, user, and keystrokes.

key (home and function keys), the researchers help him. He wanted to use the headphone to listen the NVDA explanations. He asked for disable the touchpad.

After that, the Facilitas Player (ROSAS-VILLENA *et al.*, 2015) was tested with David. The researchers explained to him all functionalities. He reported issues about navigation and the common problems founded in other websites players. David was instructed to explore all Facilitas Player functionalities and comment while navigate.

About the button's name, the `title` attribute was used in Facilitas Player. David commented that using these attribute is a problem because he did not read the button's name using the arrow keys. `title` attribute can be read only using Tab key. And depend on the person, Tab key cannot be common to use, therefore it can be confuse to understand.

David used virtual mode to navigate, for these reason he had difficulties using the Facilitas Player. He said *"when I navigate by the virtual mode the NVDA did not read button's names, he read 'button, button, button'. Then I need to navigate with Tab key to read the names. It is confuse."*

He continued testing the functionalities, but suddenly the focus disappeared from screen. He was asked about which part of the website he was located. He said that he was navigating by the hidden menus (search and configuration) and the transcript. He said *"I will need to enable that option. I was just navigating with arrow keys, then eventually lost focus in that area. Are you hidden it?"*. The answer is yes, in Facilitas is hidden that options and it is necessary to modify using `display:none`. He said *"the audio description is playing with the NVDA captions, that is cool, are you using aria-live?"*. The answer is yes.

The audio description (AD) using `aria-live` property³ was developed. These property indicates that an element will be updated, and describes the types of updates the user agents, assistive technologies, and user can expect from the live region. The `aria-live` attribute is the primary determination for the order of presentation of changes to live regions.

He was asked if he was using the audio description. He said *"yes, the audio description belongs to the player or the video?"*. The researchers explained that the audio description belongs to the player, it is a mp3 file that it was created and played along with the video. To create the mp3 file it was done through the following steps: (1) create the text that will be the audio description and put it on the caption file (`srt`) to mark the time that will be used to play the audio; (2) copy each sentence created in step 1 without the time; (3) if some word requires special pronounce, it is necessary to write as the word is spoke to guarantee the right pronunciation by the TTS; (4) use any tool to convert the text into audio, it was used Soar⁴ because is a free online tool that converts text into audio (mp3 file); and to finish, (5) it is necessary to edit the original audio and add the audio description audios in the correct time. The final mp3 file needs to have the same

³ Accessible Rich Internet Applications (WAI-ARIA) 1.0 - <http://www.w3.org/TR/wai-aria/states_and_properties>

⁴ www.soarmp3.com - Soar is a free tool to convert text to audio.

time of the video. To edit the audio it was used GoldWave⁵ editor.

When he navigated using virtual mode, he found the close button and he did not understand why need to close something that he did not open. This problem happen because the search and configuration windows were hidden back to the video using opacity. David commented *“If you use opacity does not work, because it is visible to the screen reader, you need to hidden by the display to really take away of the document flow”*. The Google Chrome browser was used at the beginning, but after 20 minutes David asked if he could use the Mozilla Firefox. He said that NVDA adapts better to Mozilla Firefox because is the browser most accessible to the people with visual disabilities.

Facilitas has help functionality. David was instructed to navigate to help to read the shortcuts. He navigated to the help button and pressed enter. The help window opens over the video, but he did not detected. He was asked if he can access the help content. He said *“I can not find the help content. You told me that a window is open over the video. The sighted people saw the dialog automatically but when we navigate, we interact with the document in a linear way, for these reason, when I pressed enter on the help button I have the sensation that nothing happen. You need to put the focus on the dialog, but also the dialog need to be marked with the correct roles. You can use JQueryUI dialog, it is very well developed. There is another ARIA property called ARIA-PopUp, you set with true value and the screen reader announce that the button has a pop-up, it is intuitive”*.

He also said *“as we are visually impaired, something that is fundamental in webpages is semantics. A signed person press the button and open a window as a dialog, for these person is clear that it is a dialog, is visible to him, by the shape, the colors, the layout, behavior, finally, the person can identify it quietly. But for people who are visually impaired, the screen reader always plays the semantic role that the html element has in the document, for example, if this dialog has been implemented with only a div, the div in the html does not have any semantic role, then for us, when we access the button, the content appears anywhere in the document, we can not tell that it is a dialog or it is separate from the document in some way or that it has a specific role in the document. For these reason, the W3C created ARIA, ARIA is actually a way to assign specific semantic roles to html components”*.

David could not read the help content, the researchers said that he can press the P key to pause and play the video. He said *“P?, interesting”*, he tested and using a lot during the rest of the test.

David was asked about the annotation functionality, he tested three annotations and said *“very interesting, very nice, it is an interesting idea. It facilitates a lot, help navigation, when you want to go to some point, you did not need to control the reproduction time, you can go by the shortcut (talking about the annotation links). This functionality is really relevant”*.

⁵ www.goldwave.com - GoldWave is a tool to edit audio.

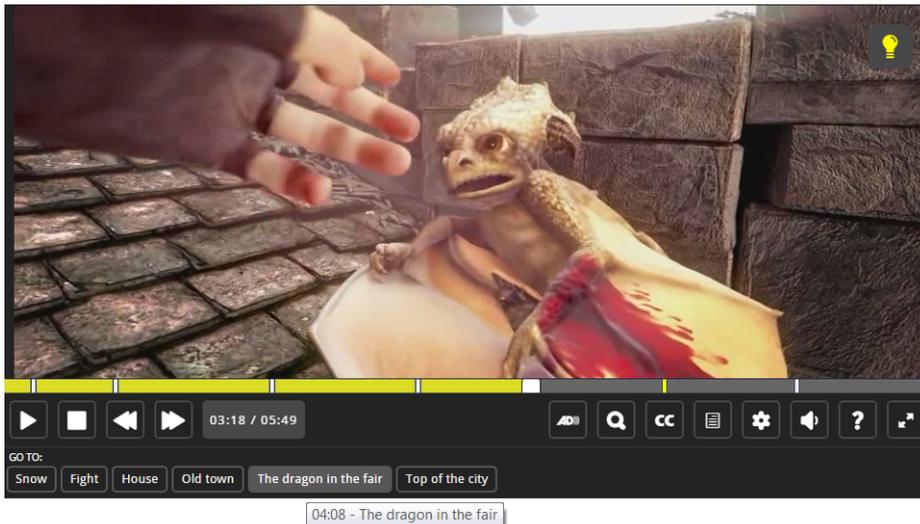


Figure 13 – Facilitas Player with the answer of the task.

David gave important contributions about how web can improve Facilitas player and how other accessible web applications can be developed for visual impaired people.

A task was prepared, which is a question about the video content. He needed to show the part of the video where the answer was, using any Facilitas player control. The first 5:49 minutes of Sintel film (BLENDER, 2010) was used, see Figure 13. The question was “The dragon had an injury. In which part of the body?”. He used NVDA search and the researchers saw that the annotation “The dragon in the fair” was selected, he accessed the annotation and played the video until it finished. Then he said that did not find the part of the video with the dragon injury. The researchers said that the scene appears before of the selected annotation. He clicked on "Old town" annotation and found the answer. The researchers asked him if he was searching using NVDA or the search functionality of Facilitas. He said “*oh yes, I was using the NVDA search, it is intuitive. I went to annotations, I could have gone to the search control and search into the caption. I did not think about that*”.

Lessons Learned

A test with David was conducted and, regarding to the observed aspects, the researchers have identified that the blind user usually navigates using virtual mode and many sites are not prepared for this type of navigation, usually they use the focus mode. A limitation of this test is that it was tested only with one user with visual disability.

As future work the researchers will improve Facilitas player functionalities, for instance, improve help functionality, facilitate navigation when the arrow keys are used and increase other functionalities, like transcript and language.

3.2 Pre-production

The Pre-production stage is important because during a video planning, it is possible to think about accessibility in ways that ensure that all members of the audience can access the video content.

At this stage, besides the activities described in the [section 2.2](#), other activities are added in order to incorporate recommendations related to accessibility. These activities are preparing the script, defining the target audience, and defining the type of video. Following, each of these activities are explained.

Preparing the script

As defined before ([section 2.2](#)), a script is composed by a set of images and audio content that are to be integrated into the video. This activity is the key moment to think about accessibility resources. Following, there is a list with recommendations during scripting:

1. **Analysis of image conditions.** [Carrera et al. \(2013\)](#) suggest an analysis of the image conditions as luminosity, contrast, and approximation to reality. This recommendation enables a clear interpretation of the visual content and avoiding a misunderstanding.
2. **Analysis of the most important content.** The most important content should be transmitted verbally (including acknowledgments and content information at the end of the video) ([DO-IT, 2013](#)). Such recommendation makes it possible for people with visual disabilities to understand the video content, because it is in audio and not in images. Furthermore, it contributes to the creation of closed captions.
3. **Analysis of captions speed.** [Carrera et al. \(2013\)](#) explained that captions cannot be too fast and suggest the identification of rate problems. In addition, case studies (see [section 3.1.1](#) and [section 3.1.2](#)) reported that participants had difficulties to read all caption because the speaker talks fast. With such recommendation, the author may consider volume of caption information transmitted in a period.
4. **Analysis of silence.** It ensures that the video's audio has enough silence to add audio description, which will be added in the authoring stage to be accessible to blind users ([DO-IT, 2013](#)). [Carrera et al. \(2013\)](#) suggest the addition of some time without sound when the scene changes, because people with visual disabilities need the description of the scenario. Formore detailed explanations, see [section 3.4.3](#).
5. **Simplicity of words.** Case studies (see [section 3.1.1](#)) reported that users have difficulties in understanding complex words used in captions. In addition, people with auditory impairments have difficulties to understand sentences with a complex structure. In this way, using short/simple words and phrases can make video content more accessible, not

only for people with auditory disabilities but also for people with cognitive and learning disabilities. [Carrera et al. \(2013\)](#) suggest avoiding unnecessary phrases or abbreviations.

Defining the target audience

The most essential activity in the product development is to understand who the users are and what their needs are. Describing the user profile is important to understand for whom the product is being constructed and help to perform future activities of analysis and evaluation ([BARBOSA; SILVA, 2010](#)).

The user profile is a description of the users' characteristics, [Barbosa and Silva \(2010\)](#) aimed to classify these characteristics according to the demographic data (age, gender, socioeconomic status), working experience, education / literacy, computer experience, an specific product experience, available technology, training, attitudes, field knowledge, goals, tasks, severity of errors, work motivation, and languages.

Therefore, the author of a video should take into consideration the wide variety of characteristics of final users in order to make the video accessible to a broad audience ([DO-IT, 2013](#)). A diversity of users, with different abilities and limitations are present on the web. Some of them have disabilities from birth or caused by an illness, disease or by an accident, or they may develop impairments with age ([ABOU-ZAHRA, 2012](#)). An example to illustrate that defining the target audience is an important activity is that users found that captions are showed faster than they can read (see the first iteration of preliminary studies, [section 3.1.1](#)). This issue can be avoided taking into account a diversity of people as audience.

Using this approach, a related literature investigation reveals that the target audience can be categorized by:

- Disabilities - Auditory, Cognitive and neurological, Physical, Speech and Visual ([ABOU-ZAHRA, 2012](#)).
- Age group - children (until 11 years old), teenager (between 12 and 18 years old), young (between 19 and 39 years old), middle-aged (between 40 and 59 years old) and older (more than 60 years old) ([W3C, 2008d](#); [LAW8069, 1990](#); [LAW10741, 2003](#)).
- Literacy - basic (language proficiency and numeric), scientific (knowledge and comprehension of the scientific concepts), economic (ability of identify economic problems, alternatives and benefits) technologic (knowledge about the technology), visual (interpret, use, enjoy, and create images and video), information and media (evaluate information through a variety of media, recognize when information is needed), multicultural (understand and appreciate the similarities and differences in habits, values, and beliefs of their own culture and others), and global (recognition and understanding of relationships

between international organizations, private and public entities, sociocultural groups and worldwide individuals) (FAJARDO; ORTENZI, 2007).

Defining the category

The video category is important as an item to be included into the transcript file (see [section 3.4.3](#)). Also, some recommendations about video player functionalities are listed aiming that the author can choose which video player is more appropriate for his/her intention.

There are several classifications of videos. For example, Facebook categorizes videos into beauty & fashion, business, cars & trucks, comedy, cute/animals, entertainment, family, food & health, home, lifestyle, music, news, politics, science, sports, technology, and video gaming. YouTube categorizes videos into film & animation, autos & vehicles, music, pets & animals, sports, travel & events, gaming, people & blogs, comedy, entertainment, news & politics, howto & style, education, science & technology, and nonprofits & activism.

Preliminary studies (reported in the [subsection 3.1.1](#), [subsection 3.1.2](#) and [subsection 3.1.3](#)) performed by the author of this thesis showed the five categories that users (students, adults and older people) chose as the most interesting to watch a video. Some accessibility recommendations according to category are listed as follows:

- **Entertainment** (music, films, TV show, sports, soap operas, anime, etc.) - Users usually watch entertainment videos in other languages, therefore it is important to provide closed captions as alternative content and also provide other languages for caption. People who watch videos with captions, should be able to configure their preferences, for example, change the font color, font size, font family, and caption background.
- **Information** (news, documentaries, science, technology, health, etc.) - Users usually watch videos with a lot of information. Localizing some part of the video can be an arduous task when a video is long, it is important to provide a search functionality help with this task. Nevertheless, it is necessary provide closed caption for searching. Closed captioning helps users with auditory disabilities to understand the content because if they did not listen some word, they are able to read. When users watch long videos, it is important to provide annotations or chapters to jump to specific content, avoiding the waste of time.
- **Learn new things** (cooking recipes, makeup, languages learning, tutorials, how to do videos, etc.) - Users often need to review past excerpts from a video to reinforce learning or to make connections between the information presented. Therefore, features like rewind and forward are essential. However, these can be tedious to use when the searched content is far from the current video instant. Thus, search tools are good options. The search can usually be performed by keyword, analyzing the closed caption. Then, the presence of closed captioning in videos is an important aid.

- **Education** (teach some concept, create an autobiography) - Teachers usually show excerpts from videos to their students to improve or better explain some concept. Therefore, the option to mark specific parts of the video is useful, because they avoid annotating by hand the video instant that they want to show. Teachers use closed captions to help students with the grammar, they can reinforce how a word is spelled or listened.
- **Things of everyday life** (pet videos, babies, etc.) - Closed caption is an alternative content that helps many users to understand the video content, including people with disabilities. Another alternative content to help the understanding is the transcript.

3.3 Production

The production stage consists of recording the video. [section 2.2](#) described two types of video capture: spontaneous and ad-hoc (planned) capture. Some recommendations for these types of capture are in the following paragraphs.

Recommendations to spontaneous capture

- 1. Quality.** Mobile devices are associated to this type of capture, sometimes authors use the resolution of the application installed in the mobile device (YouTube, Facebook, WhatsApp, Instagram, etc.), and other times they choose the resolution. It is important to capture with the best resolution ([KIRK et al., 2007](#)), because the lack of quality could affect final users, specially users with visual disabilities as they may not visualize the image of the video correctly.
- 2. Actors talking at once.** Authors have difficulties to create captions because they did not know the order of the phrases (according to the case studies results, see [chapter 4](#)). It is recommended that each actor talk one at a time to make captioning easier.
- 3. Colloquial language.** Authors explained that they should provide alternative content when actors are using a colloquial language to facilitate the final users understanding (according the case studies results, see [chapter 4](#)).

Recommendations to ad-hoc capture

- 1. Quality.** Participants reported that low resolution on videos is an issue when they want to watch as full screen (see [section 3.1.2](#)). It is important to choose the best resolution setting of the device for recording video. At the post-production stage, it is possible to generate videos with different resolutions without significant loss of quality.
- 2. Text into images.** It is important not to use text into images because they are hard to be rendered as text and, because of that, they cannot be easily read by automatic tools like

screen readers. Therefore, it is necessary to minimize the texts for display and clearly indicate them in the script to plan an accessible alternative (CARRERA *et al.*, 2013).

- 3. Volume.** Older people with auditory impairments said that low volume is common when they watch movies (see [section 3.1.1](#)). It is essential that authors record in a high quality sound and with no background noise that can muffle the voice (CARRERA *et al.*, 2013).
- 4. Environment sound.** Carrera *et al.* (2013) recommended the recording of environment sound, because it helps people with visual disabilities or blindness locate the action. Therefore, it is recommended to record the sound environment a time before and after of the scene, that will help to include audio descriptions.
- 5. Caption space.** Ensure that photographers are careful during filming to include adequate screen space for captions. Consider affixing duct tape to the bottom of their viewing screens to prevent them from filming key details in the area where captions will appear (DO-IT, 2013).
- 6. Record the face.** In an interview, people with moderate hearing impairments could benefit from the recording of the face of the character, because they can read the lips, reducing, then, some people's difficulties (CARRERA *et al.*, 2013).
- 7. Avoid people talking all at once.** It is important to take into consideration the producer intention. If the producer wants to spread the message of actors talking all at once and the listeners did not understand, authors should caption as an additional information explaining the action, for example, [people talking about the weather]. Otherwise, it is recommended that each actor talk one at a time.

3.4 Post-production

The Post-production stage starts when the media is obtained, either from the author him/herself (original video), or from other authors (third parties videos). When the media is from other authors, it is necessary to read the copyright or request authorization.

[section 2.2](#) described the video editing was described as the main activity of post-production stage. In this section, two activities are added: the video validation to identify seizure risks and the method for alternative content authoring called Video4All proposed in this thesis.

3.4.1 Video editing

During the video editing it, is important to provide characteristics to make the video flexible for different devices and for people with diverse abilities. These characteristics are following explained.

- **Add alternative content to text.** If the text is important, for example, when it includes titles, credits and acknowledgments, it is better to add an alternative content to the text contained in the video. In this case, a narrator reading the text allows people with visual disabilities to understand the content.
- **Resolution.** Create video copies with different resolutions. When a low resolution video is viewed as full screen mode, the visualization to final users is affected, especially for users with low vision. It is recommended to provide high quality videos. The video delivery system can choose the resolution that best fits the final user needs.
- **Video file.** A video file consists of a container format (e.g. ogg) containing video data in a video coding format (e.g. Theora) alongside audio data in an audio coding format (e.g. Vorbis). There are different container formats, the most frequent video file type are: .wmv (Windows Media Video), .mpg, .avi (Audio Video Interleave), .mp4, .ogv (which belongs to the open container format OGG), .mov (formatted from Apple), WebM, etc. However, current implementations of different Web browsers do not support the same media formats in their implementations of HTML5 video, therefore, it is important to provide at least two types of formats to ensure the video will be played. There are 3 supported video formats for the <video> element of HTML5: MP4, WebM, and Ogg ([W3SCHOOLS, 2015](#)).
- **Words Per Minute (WPM).** The quantity of words per minute influence on how fast the captions are. Case studies conducted in the present study reported that participants had difficulties to read all caption because the speaker talks fast (see [section 3.1.1](#) and [section 3.1.2](#)). This issue may be caused due to the quantity of words per minute not being taken into account. The Described and Captioned Media Program (DCMP)⁶, provides recommendations about the ideal WPM rate which states that a video should have: All lower-level educational media should optimally be captioned at a presentation rate which does not exceed 130 words per minute (wpm). Middle-level should not exceed 140 wpm. Finally, upper-level are not to exceed 160 wpm. The only occasion when presentation rate is ignored is when any person is quoting a well-known or famous person speaking on-screen or off-screen. The same way when poems and other published works are quoted, and/or song lyrics are sung. These must be captioned verbatim.

3.4.2 Video evaluation to identify seizure risks

People who present seizure disorders (including different types of epilepsy and migraines) may react to visual flickering or audio signals at certain frequencies or patterns ([ABOU-ZAHRA, 2012](#)). For this reason, video content should be evaluated to avoid any risks. As a general rule, a video should not contain anything that flashes more than three times in any one second period ([W3C, 2008e](#)).

⁶ Presentation rate <http://www.captioningkey.org/presentation_rate.html>

Before continuing with the video authoring process, it is important that the video evaluation be performed to identify seizure risks. Tools can help authors to determine whether the video in their content is likely to cause seizures. If the video presents flash failures, it is necessary to go back to video editing and remove those frames.

The Trace Center's Photosensitive Epilepsy Analysis Tool (PEAT⁷) is a resource for developers to identify seizure risks in their web content and software. The evaluation used by PEAT is based on an analysis engine developed specifically for web and computer applications (CENTER'S, 2015).

3.4.3 Video4All - Method for alternative content authoring

Currently, common users are Web content producers, specially video content producers. They record videos using cellphones, post videos on the Internet and share online video on web sites. However, they have difficulties to produce accessible videos because there is a lack of tools and methods to produce alternative content.

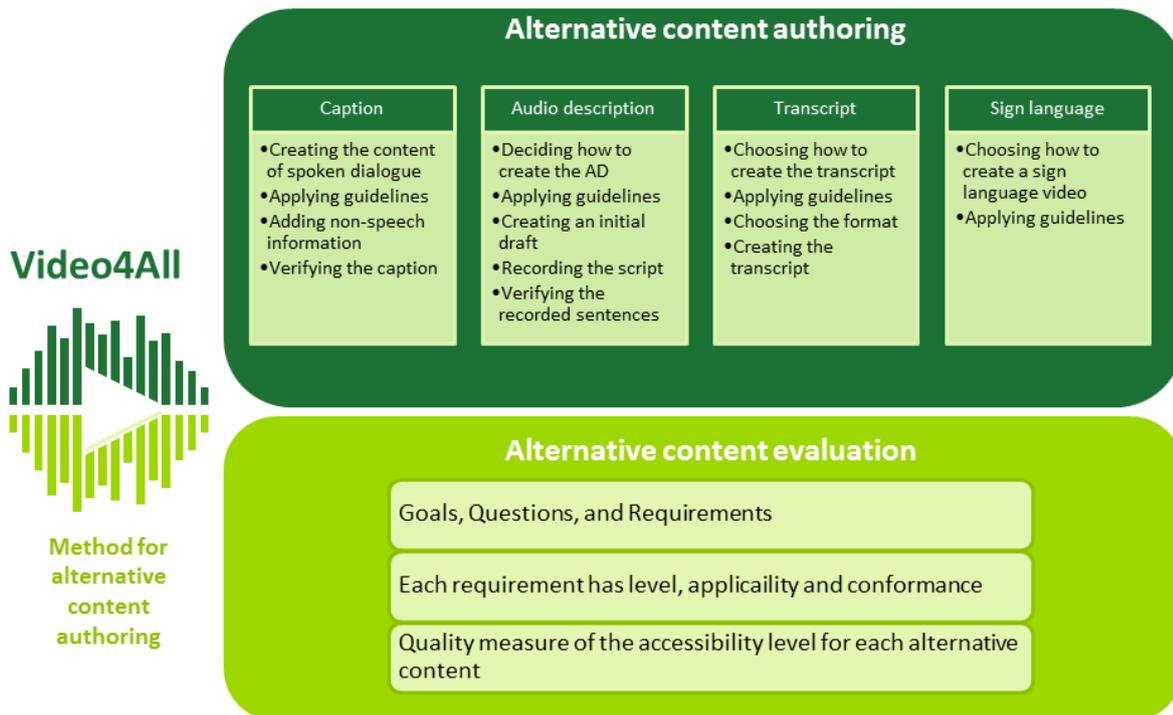
As defined in section 2.1, alternative content is a Web content that serves the same function or purpose as the original content. The method for alternative content authoring, called Video4All, is composed by a set of activities for alternative content authoring and alternative content evaluation. The Video4All is shown in the Figure 14. In this figure, the alternative content authoring is divided into four alternative content: caption, audio description, transcript and sign language. Each alternative content is composed by activities. For caption authoring, it was defined four activities: creating the content of spoken dialogue, applying guidelines, adding non-speech information, and verifying the caption. For audio description authoring, it was defined five activities: deciding how to create the AD, applying guidelines, creating an initial draft, recording the script, and verifying the recorded sentences. For transcript authoring, it was defined four activities: choosing how to create the transcript, applying recommendations, choosing the format, and creating the transcript. Finally, for sign language authoring, it was defined two activities: choosing how to create a sign language video and applying guidelines for sign language video. The alternative content evaluation is composed by: goals, questions and requirements, each requirement has level, applicability and conformance, and quality measure of the accessibility level for each alternative content.

Alternative content authoring

Every person should have the opportunity to contribute with the production of alternative content to provide access to the video content for people with disabilities. The alternative content authoring describes activities for alternative content. Four types of alternative content were selected (as defined at section 2.1): captions, audio description, transcript and sign language. The activities were formalized after the case study with authors (see chapter 4). During the case study,

⁷ PEAT is available for downloads on <<http://trace.wisc.edu/peat/>>

Figure 14 – Video4All - Method for alternative content authoring



Source: Elaborated by the author.

authors were instructed to create alternative content (captions, audio description and transcript). They were not instructed to create sign language because it would require time to learn a new language. They were supported by a Website⁸ where for each alternative content, it was described the concepts, benefits, steps for authoring, guidelines for creation and validation. Authors can choose the information to read to accomplish the task (creation of alternative content). They did not read all guidelines, they selected it according to the video content, for this reason guidelines were grouped by a common goal. The following sections describe all activities for authoring for each type of alternative content.

Captions

In order to understand the needs behind the use of closed caption, two related experiments were conducted, described at sections 3.1.2 and 3.1.1). The first one shows an evidence that older people find it harder to understand video content, therefore all of them watched the videos with captions. The second one shows an evidence that participants (8 students and 2 older people) want to produce alternative content, they expressed that they would like to edit the captions.

Every caption must comply the following characteristics of quality (DCMP, 2014):

- **Accurate.** Errorless transcript is the goal for each production.

⁸ <<https://sites.google.com/site/videoaccessivel>>

- **Consistent.** Uniformity in style and presentation of all captioning features is crucial for viewer understanding.
- **Clear.** A complete textual representation of the audio, including speaker's identification and non-speech information, provides clarity.
- **Readable.** Captions are displayed with enough time to be completely read, they are synchronized with the audio, and not obscured by (nor do they obscure) the visual content.
- **Equal.** Equal access requires that the meaning and intention of the material is completely preserved.

For caption authoring, the method defined four activities: creating the content of spoken dialogue, applying guidelines, adding non-speech information, and verifying the caption.

The **first activity** to start captioning is **creating the content of spoken dialogue**; this content is synchronized with the audio track. In other words, the spoken dialogue is divided into phrases and these phrases appear on screen at the same time as the spoken words. There are five ways to obtain the spoken dialogue:

1. *Using the video's script.* If the video was planned, the script provides specific information about the audio track content that can be used to help with the creation of spoken dialogue.
2. *Using the transcript.* If the transcript was created, it provides all dialogues and non-speech sounds that can be used with the creation of spoken dialogue and non-speech information.
3. *Using a voice-to-speech recognition software to get a rough transcript.* There are programs as Dragon Speech Recognition Software⁹ or YouTube's speech recognition technology. Since captions are automatically generated, the author needs to review them.
4. *Manually.* The author watches the video and writes everything that is said. There are free captioning and editing programs which help on this process, such as Aegisub¹⁰, Amara¹¹ and Subtitle Horse¹². In addition, there are paid tools such as MAGpie¹³, Dotsub¹⁴ and CaptionKeeper¹⁵.
5. *Paying a third party.* The Described and Captioned Media Program (DCMP¹⁶) provides a listing of captioning vendors¹⁷.

⁹ Dragon Speech Recognition Software is available for purchase on <<http://www.nuance.com/dragon/index.htm>>

¹⁰ Aegisub is available for download on <<http://www.aegisub.org/>>

¹¹ Amara is available online on <<http://amara.org/>>

¹² Subtitle Horse is available online on <<http://subtitlehorse.com/>>

¹³ MAGpie is available for purchase on <http://ncam.wgbh.org/invent_build/web_multimedia/tools-guidelines/magpie>

¹⁴ Dotsub is available for purchase on <<https://dotsub.com/>>

¹⁵ CaptionKeeper is available for purchase on <<http://ncam.wgbh.org/webaccess/captionkeeper/index.html>>

¹⁶ The Described and Captioned Media Program is available on <<https://www.dcmp.org/>>

¹⁷ Captioning Service Vendors are available on <<https://dcmp.org/ai/10/>>

People who choose captioning by themselves can continue with the following activities. The **second activity** is **applying guidelines**. The DCMP Captioning Key developed guidelines to create closed captions explained on [section B.1](#), [section B.2](#), [section B.3](#) and [section B.7](#).

The **third activity** is **adding non-speech information** such as speaker identification, music, laughter and meaningful sound effects. To add the speaker identification, two different semantics was found in the literature. It is necessary to choose one type and be consistent throughout the production.

1. Capitalize proper nouns for speaker identification. All other speaker identification should be lowercased unless this identification is being used as a proper noun (DCMP, 2014). Examples: (Tiago) (female narrator). Note that they use parentheses to indicate the name. Guidelines for speaker's identification are described on [section B.5](#).
2. To denote speaker identification, capitalize proper nouns and use colon at the end, for example, Mr. Kishi: (CAB, 2012).

In addition, two semantics for non-speech information are found in the literature. It is necessary to choose one type and be consistent throughout the production.

1. A description of sound effects, in brackets, should include the source of the sound, for example, [audience cheering]. If a description is used for offscreen sound effects, it is not necessary to repeat the source of the sound if it is making the same sound a few captions later, for example: [squealing continues]. Offscreen sound effects should be italicized, if italics are available. It includes background music. For example, [*thunder rumbling*] (DCMP, 2014). Guidelines for non-speech information are described on [section B.4](#) and [section B.6](#).
2. To denote descriptions of non-speech information use parentheses and uppercase the first letter (Audience laughter) (CAB, 2012).

Finally, the last activity is **verifying the caption**, which consist of reviewing the caption created. The author should verify if the caption is synchronized with the video content. Then, the author should listen all audio of the video and check if the caption appear at the correct time. The author should review the grammar and punctuation of the caption, verify if all significant sounds were described, and check if all speakers were identified.

Audio description

An audio description (AD) needs to be provided when a video contains visual elements that must be observed in order to understand it. AD is indispensable for blind people to understand the visual content presented on the video. As a research result (ROSAS-VILLENA *et al.*, 2015)

(video accessibility issues for people with visual disabilities), a blind participant remarked that he never found videos with AD when he watches videos.

For audio description authoring, the proposed method defined five activities: deciding how to create the AD, applying guidelines, creating an initial draft, recording the script, and verifying the recorded sentences.

The **first activity** to start with the audio description authoring is **deciding how to create the AD**. There are four ways to add AD to videos:

1. *Building the AD into the video*. Each character identify himself and his function on the video, so that any people know who is speaking (DIGITALGOV, 2014b). For example, “Hi, my name is Victor, I’m secretary of sports club”.
2. *Making another video*. When the video player does not support the option to turn AD on and off, create two versions of the video: one with audio descriptions and one without. For this, a copy of the original video needs to be done, in which the AD track is integrated with the original audio track (DIGITALGOV, 2014b). For example, for videos with AD, the labels could be “Title of the video with audio description”. For the same video without AD, it could be “Title of the video without audio description.”
3. *Adding an extra audio track*. When the video player supports more than one audio track and the user can turn AD on and off (DIGITALGOV, 2014b). Then, is necessary to record a separate audio track that would be played with the video.
4. *Paying a third party*. The Audio Description Project elaborated a list ¹⁸ with some organizations that provide describers.

The **second activity** is **applying guidelines**. The Audio Description Coalition of the United States, developed standards for audio description. These guidelines are described on [Appendix C](#).

People who choose to create their own AD can continue with the following activities.

The **third activity** is **creating an initial draft** of AD.

1. *Using the video’s script*. If the video was planned, the script includes details on any graphic required that can be used to describe visual information.
2. *Using the transcript*. If the transcript was created, it provide significant visual information. This information could be described in long sentences, thus it would be necessary to adapt it according to the space of time available.

¹⁸ [<http://www.acb.org/adp/adrates.html>](http://www.acb.org/adp/adrates.html)

3. *Manually*. This option is for doing the transcript yourself. Guidelines to help with this process were explained in the last activity.

The **fourth activity** is **recording the script**. This activity provides an initial audio description, this audio will be tested with the original audio in the next activity. There are two ways to do this activity:

- *Record a narration performed by a human*. Pay a narrator to record or made yourself. Recorded by a narrator is best for users because humans can speak with the correct intonation and emotion, but the cost is high.
- *Use a speech synthesizer program to convert text-to-speech*. There are tools to help with this step, for example Soar¹⁹. If the script has any word in another language, write the pronunciation in the recording language.

The **fifth activity** is **verifying the recorded sentences**.

It is important to verify if each sentence fits into the silence chunks of the original audio track. If not, cut some words or record again. The audio description cannot overlap the original audio because it can cause misunderstanding of both audios. GoldWave²⁰ and VideoMelt²¹ are two examples of software that can help in this task. The first one is available only to audio editing while the other can edit audio, video and images.

Also, it is important to verify the video player requirements. Some video players require the audio track to have the same duration as the video. Thus, it is necessary to add silence between the sentences to guarantee that each sentence will play in the correct time. In others video players, the AD is embedded into the video, thus it is necessary to incorporate the AD into the original audio track.

Transcript

A transcript should contain the dialogs, significant visual information and non-speech sounds. For transcript authoring, the method defined four activities: choosing how to create the transcript, applying recommendations, choosing the format, and creating the transcript.

The **first activity** to start the transcript is **choosing how to create the transcript**. There are four ways for creating the transcript:

1. *Paying a third party*. There are transcript services that transcribe audio files and provide the transcript in various formats. Costs depend on quality and turn-around time, which

¹⁹ Soar is a tool to generate audio from text and have only two languages for conversion, English and Portuguese, available on <https://www.soarmp3.com/>

²⁰ Goldwave is available for download on <http://www.goldwave.com/>

²¹ VideoMeld is available for download on <http://videomeld.com/>

ranges from hours to weeks (LAWTON, 2009). A list with transcript services are at uiAccess website²², providing the transcript in languages as English, French, German, Italian, Japanese, Portuguese, Romanian, Russian, Spanish and Turkish. There are many options for English transcripts, but for any other. For example, for Portuguese only two options are offered, the cost starting at \$1,25 per minute for up to 2 speakers, or \$4.50 per minute for more than 2 speakers. The uiAccess recommends avoiding user editing effort, provide the transcriptionist a list of names, acronyms, abbreviations, technical words, jargon, and anything else they might not recognize in the audio (LAWTON, 2009).

2. *Using speech recognition software.* Speech recognition software helps to have a draft of the transcript, it is necessary review to correct mistakes. Some examples of these software are described below:

- Dragon NaturallySpeaking²³, might be a viable option if user has a lot of media with a single speaker, such as a regular podcast that is mostly have one person. Such software requires "training" for a particular voice, so if your audio is interviews with different people, this will not work as well (LAWTON, 2009).
- Transcribe²⁴ is an application that allow audio to text conversion, which costs US \$20 for an annual license. It is possible to use the trial version for a week.
- Express Scribe Transcription Software²⁵ has the option to convert speech to text, but only with a paid version.
- YouTube provides speech recognition technology to automatically make captions available with some restrictions (the video is too long or has poor sound quality, there are multiple speakers whose speech overlaps, there is a long period of silence at the beginning of the video and the language in the video is not supported). It is possible to edit the captions or remove them from the video.

3. *Manually.* Type the transcript yourself. Doing it yourself is likely to be frustrating. Transcripts for audio podcast could take at least three times as long as the audio to type it up, half an hour for a 10-minute podcast (LAWTON, 2009). And according the case studies with authors (see chapter 4), transcripts for videos could take five times as long as the video cross-out, because visual information is also described. There are free software products that can help by slowing down the text and providing easy pause buttons, for example, Express Scribe Transcription Playback Software²⁶.

4. *Using the captions and audio description.* If captions and/or audio description are available for the video, using these information makes it easier to do the transcript. Captions contain

²² Transcript services - <http://www.uiaccess.com/transcripts/transcript_services.html>

²³ Dragon NaturallySpeaking is available on <<http://www.nuance.com/dragon/index.htm>>

²⁴ Transcribe is available on <<https://transcribe.wreally.com/>>

²⁵ Express Scribe Transcription Software is available on <<http://www.nch.com.au/scribe/>>

²⁶ Express Scribe Transcription Playback Software is available on <<http://www.nch.com.au/scribe/>>

the dialogue and non-speech sounds, and audio description contains visual information. All these pieces of information are needed to do a transcript, but also, it is possible include more detailed information. For example, AD includes information that fit in the silences, in the transcript these information could be extended because the time is not a limitation.

The **second activity is applying recommendations** based on [Lawton \(2009\)](#), [W3C \(2008a\)](#) and [DigitalGov \(2013\)](#). Only these recommendations were found on the Web, guidelines for transcript were not found. Below are described recommendations to create transcripts:

1. Include speakers names. [Lawton \(2009\)](#) suggest some recommendations:
 - Capitalize proper nouns and use colon at the end, for example, “John: Hello.” The W3C in their transcripts uses bold for names, for example, “**John**: Hello.” When is technically possible, use bold.
 - For an informal podcast transcript with multiple speakers, it’s probably best to use speakers’ full names the first time, and then just their first names.
 - When there are multiple speakers, use hanging indents to make it easy to skim for a particular speaker. It is possible use CSS and give each speaker a different color or style. (Make sure to include names so there is redundant color coding.)
 - When the focus should be on the interviewee’s answers and not the interviewer, format the interviewers questions in smaller, lighter text so the interviewee’s answers stand out more clearly.
2. Include all speech content. If there is speech that is not relevant, it is usually best to indicate that it has been excluded from the transcript. All additional information should be in brackets. Example: “[participants discuss the weather while the presenter reboots his computer]” ([LAWTON, 2009](#)).
3. Include relevant non-speech audio ([LAWTON, 2009](#)). Non-speech audio is related with the use of irony, music and scenery. It is important describe only the relevant sound. Example “[rose petals were thrown from a helicopter]”. The irrelevant background noise is generally not described.
4. Include relevant visual information ([W3C, 2008a](#)). Visual information include visual context, actions and expressions of actors, and any other relevant visual information (photographies, unusual dress). For example “[photography of a woman with a blue macaw inside a picture frame that says best friends],” “Joe: I hate this computer! [shouted].” [Lawton \(2009\)](#) explained some recommendations in presentations. First, instead of pointing to a slide and describe, “as you can see on this slide, the traffic peaked here”, describe, “this chart of website traffic for the last year shows that it peaked in August.” Second, if the presenter does show-of-hands (e.g., “How many people follow WCAG 2.0?”), describe

the results for the audio recording (e.g., “about half”). And third, repeat questions that are not picked up by the audio recording.

5. Check the transcripts for proper punctuation (DIGITALGOV, 2013).
6. If the person speaking uses any acronyms, define them (DIGITALGOV, 2013).
7. Edit according to situation. Lawton (2009) proposed some directions:
 - In some cases, such as legal depositions, the transcript must be verbatim, including ums, ahs, and indicating pauses.
 - For most podcasts, presentations, and such, minor edits for readability are appropriate. For example, it would generally be correct to edit: "The success criterion in WCAG 1.0 - I mean in WCAG 2.0 - are more clear and testable." to: "The success criterion in WCAG 2.0 are more clear and testable."
 - It is not appropriate to change the meaning from the audio. Likewise, it is usually not appropriate to significantly correct grammar or other mistakes.
8. Add navigation and clarifications. Lawton (2009) suggested the following recommendations:
 - Add headings and links where it will make the transcript more usable. For example, add links when the video is short and explains many concepts, create links for these concepts; or add titles when the video is long.
 - Add clarifying information, as long as it is clear that it is not part of the actual audio, e.g., words added to a paragraph put in [brackets], or separate sections with headings "Introduction", "Transcript", "Resources".

The **third activity is choosing the format**. The objective is choosing which file format the transcript will be available. In the list of transcript services (explained in the first activity), it is presented more than 20 formats, but the format most commonly used is HTML, followed by DOC. Lawton (2009) recommended to provide the transcript in HTML to facilitate the online search. According to this, in this thesis also recommended the HTML format.

The **fourth activity is creating the transcript**. Based on the examples of transcript available on the W3C website²⁷, NASA²⁸, and YouTube tips²⁹, a transcript template was developed. This template helps authors develop the transcript from a default structure. The structure is composed by three parts: the video information, the links to the video material, and the video description. This structure was defined from the examples of transcript explained before. It is

²⁷ Podcast transcript <<http://www.w3.org/WAI/highlights/200606wcag2interview.html>> and presentation transcript <<http://www.w3.org/WAI/highlights/200706wcag2pres>>

²⁸ NASA audio transcript <<http://www.nasa.gov/news/media/audiofile/03132013.html>>

²⁹ YouTube tips <<https://support.google.com/youtube/answer/2734799>>

a standardization proposal because a template for transcript was not encountered in the search performed in this study. For an example of the template, see [Appendix C](#). The transcript's template is structured as follows:

- 1. Video information.** The video information includes the type of video content, title, data (recording data, if it does not exist will be replaced by the publication data), and abstract. Type of video or category (see [section 3.2](#)) is presentation, podcast, documentary, cooking recipes, etc. For example, "Presentation: Video accessibility, February 8th 2016." An abstract is a short explanation of video content, it also could contain where the video was recorded, and who the main actors are (in case of presentations, who is presenting).
- 2. Links to the video material.** If the webpage of the transcript is different from the webpage of the video, the video webpage should have a link to transcript webpage and also the transcript webpage should have a link to the video webpage. For example, a video's page has the link "A transcript of Johana's presentation is available on the ICMC website" and the transcript's page has the link "Video with audio on the ICMC theatre". The transcript also should include the link to resources download. Resources as video, audio or slides should be available for download. For example, the link "Downloadable version of video with audio (mp4)."
- 3. Video description.** In the W3C examples, the addition of a title "Transcript" before start the transcript is common, therefore it is recommended to add this title to separate the video information and material from the video description. Also it was observed on the examples that it is usually, use a blank line to force the start of a new description and separate paragraphs by actors. If the video has more than one actor, the paragraphs are separated by actor. Each paragraph should contain all speech of that actor.

Sign Language

For sign language authoring, the method defined two activities, choosing how to create a sign language video and applying guidelines for sign language video. The method does not fully explain steps to create sign language video because as any language, sign language also requires a specific method and time to learn.

The first activity is choosing how to create a sign language video.

- *Paying a third party.* There are services that provide the sign language video of the video content. In the Unites States, there are many services to help deaf people, for example the Video Relay Service (VRS)³⁰, American Sign Language (ASL) Services³¹, LanguageLine

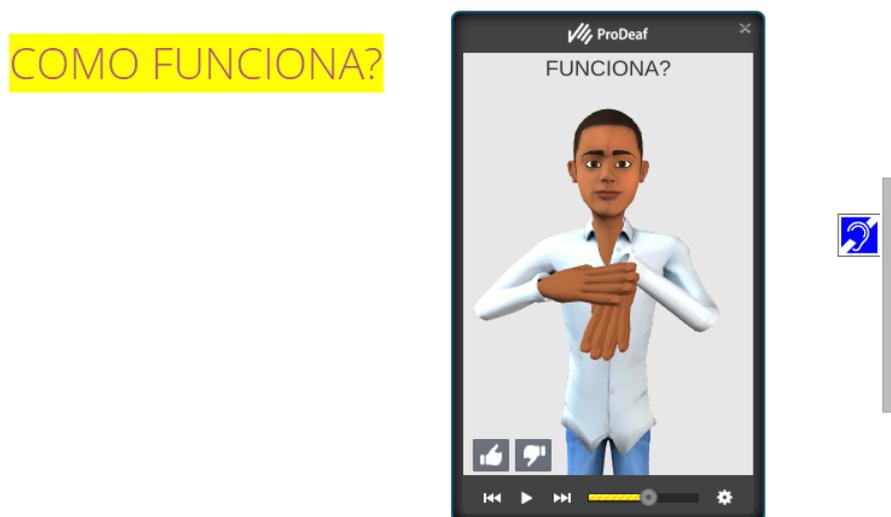
³⁰ VRS - <http://www.adapacific.org/resources/relay/vrs.php>

³¹ ASL - <http://aslservices.com/>

Solutions³². These services allow people who use American Sign Language to be able to place phone calls by signing instead of typing, video remote interpreting, onsite interpreting, personal interpreter, among others. In Brazil, there are also websites that provide some services, for example, the association of professionals and interpreters of Libras of Rio de Janeiro (APILRJ)³³, and a list with the prices for interpretation are on Sintra website³⁴.

- *Recording yourself.* This option is for people that have knowledge about sign language as well as experience. To learn sign language there are specific courses, for example the National Federation for Education and Deaf Integration of São Paulo (FENEIS-SP)³⁵.
- *Using an online translator.* There are online translators that use avatars. For example, ProDeaf WebLibras³⁶ is an online translator from text in Portuguese to Libras. Figure 15 shows the avatar when the icon next to scroll bar is clicked, it reads the Website content. Hand Talk³⁷ allows the translation from text and audio to Sign Language and also provides a solution for websites and an application for mobile.

Figure 15 – ProDeaf Web Libras avatar



Source: Adapted from ProDeaf (2015).

The **second activity** is **applying guidelines for sign language video**. The goal of this activity is informative, because amateur authors cannot create sign language videos, because is a language that require time to learn. As the sign language varies according to the country, also guidelines for specific countries are developed. For example, in the Measured Progress Innovation Lab and Maryland State of the United States produced guidelines for the development

³² LanguageLine Solutions - <<http://www.language.com/>>

³³ APILRJ - <<http://www.apilrj.org.br/>>

³⁴ Sintra - <http://www.sintra.org.br/site/pages/valores_libras_sup.htm>

³⁵ <<http://www.feneissp.org.br/>>

³⁶ ProDeaf WebLibras is available on <<http://www.weblibras.com.br/>>

³⁷ Hand Talk is available on <<http://www.handtalk.me/app>>

of American Sign Language (ASL) (LAB; STATE, 2015). In Brazil, there is a guide with explanations about auditory disabilities, laws and guidelines of Brazilian Sign Language (Libras) (UNIFENAS, 2010) or some guidelines for sign language applied to TV (ABNT, 2005). Each country name their Sign Language, for example, Spanish Sign Language (LSE), Greek Sign Language (GSL) and Indian Sign Language (ISL). Below are described guidelines for sign language according to the Brazilian National Standards Organization (ABNT) (ABNT, 2005):

1. The place where the interpreter record should have:
 - Enough space so the interpreter will not get stuck to the bottom, avoiding the appearance of shadows.
 - Adequate and appropriate lighting so that the camcorder can capture, with quality, the interpreter and the bottom.
 - The video camera should be supported or fixed on fixed tripod.
 - Ground markings to delimit the interpreter space for movement.
2. When the interpreter image is shown:
 - The height of the window should be at least half of the TV screen height.
 - The width of the window should occupy at least a fourth of the TV screen width.
 - Whenever it is possible, the window should be located so as not to be covered by black stripe of the closed caption.
 - When the displacement is needed on the TV screen, the window image should continue on the screen.
3. For a good visualization of interpretation, the following conditions must be met:
 - The clothing, skin and hair of the interpreter should be contrasting between themselves and the bottom. Background and dressing similar to the interpreter skin tone should be avoided.
 - In the transmission of TV news and other programs, with the interpreter on the scene, measures for good visualization should be taken.
 - Any other image should not be included or overlapping on the window.

Alternative content evaluation

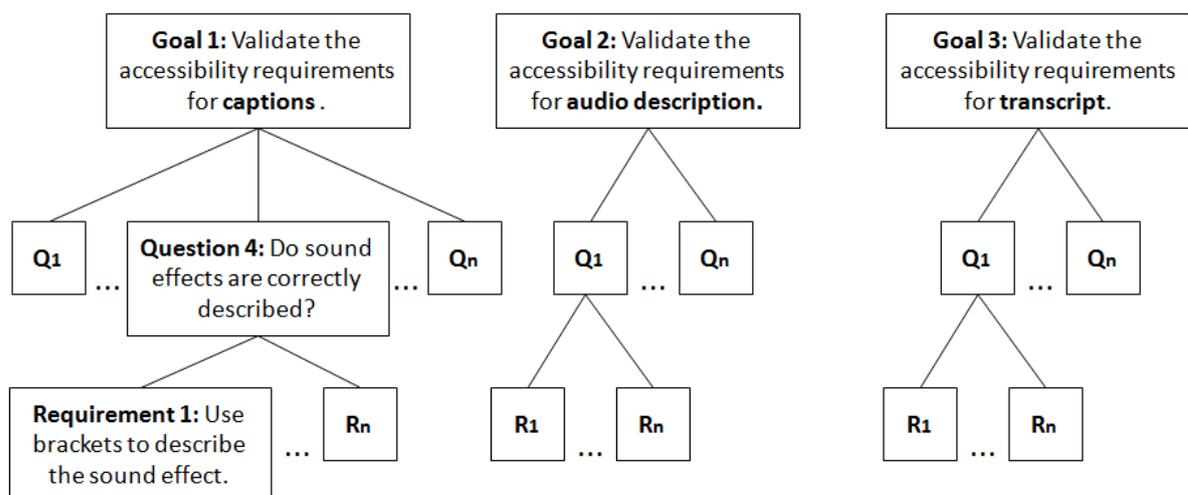
To evaluate the alternative content created by amateur authors, the method defined an alternative content evaluation with accessibility requirements to assess video. The objective of this evaluation is obtain a quality measure of the accessibility level of each alternative content. For that, it is important to take into account the accessibility requirements addressed by each alternative content and also define a classification to measure the quality of accessibility.

A method used for quantitative analysis is Goal-Question-Metric (GQM) (BASILI; WEISS, 1984). This method establishes the goals of the data collection, develops a list of questions of interest, and the metrics for each question. In GQM method, a set of metrics are defined to answer the questions in a measurable way. These metrics are collected, evaluated and analyzed in real time to provide a quantitative information to answer the questions.

Vigo, Brajnik and Connor (2012) explained that in the web engineering domain, a metric is a procedure for measuring a property of a web page or website. A metric can be the number of links, the size in KB of a HTML file, the number of users that click on a certain link, or the perceived ease of use of a web page. In the domain of web accessibility, a metric can measure the following qualities: the number of pictures without an alt attribute, the number of Level A and AA success criterion violations, the number of possible failure points where accessibility issues can potentially happen (such as the number of images in a page), the severity of an accessibility barrier, and the time taken to conduct a task. They also explain that in Web accessibility is common consider whether a web page/website is according to a set of requirements such as those defined by WCAG 2.0 or by Section 508. The metrics which are based on whether success criterion of given guidelines are met are called conformance-based metrics.

As the alternative content is composed by a set of accessibility requirements, the alternative content evaluation is inspired on GQM model, using accessibility requirements for each alternative content instead of using metrics. Figure 16 shows the GQM adapted using requirements instead of metrics. This evaluation is composed by three goals: validating the accessibility requirements for captions, validating the accessibility requirements for audio description, and validating the accessibility requirements for transcript. Each goal is composed by questions, and each question is composed by accessibility requirements.

Figure 16 – Alternative content evaluation



Source: Elaborated by the author.

The accessibility requirements were developed from the guidelines and recommendations

of the alternative content explained before at Alternative content authoring. The requirements related to functionalities of video player were excluded. After excluding, the proposed method included 109 requirements for captions, 75 requirements for audio description and 28 requirements for transcript. Tables 7, 8, and 9 show each question and the amount of accessibility requirement for captions, audio description and transcript.

Table 7 – Questions for Captions evaluation

Nº	Questions for captions	Amount of req.
Q1	Do all text (case, font, line division, caption duration, and caption placement) have a correct appearance and presentation?	12
Q2	Do captions incorporate the proper use of spelling, grammar, punctuation, capitalization, and other factors deemed necessary for high-quality captioned media?	25
Q3	Is the presentation rate adequate?	5
Q4	Are sound effects (music, narration, or dialogue) correctly described?	20
Q5	Is speaker's identification correctly specified?	8
Q6	Are captions correctly synchronized?	2
Q7	Are the intonation, play on words and no audio considered?	4
Q8	Are foreign language, dialect, slang and phonetics considered?	6
Q9	Are numbers (dates, time, periods of time, fractions, percentages, amounts, and measurements) correctly described?	27
	Total of requirements	109

Table 8 – Questions for audio description evaluation

Nº	Questions for audio description	Amount of req.
Q1	Does description describe what you see?	9
Q2	Does description describe objectively?	6
Q3	Is the description allowing the listeners to hear the dialogue?	5
Q4	Does description allow listeners to comprehend the material?	3
Q5	Does description describe the material without censorship?	1
Q6	Is language consistent?	12
Q7	Does description describe race, ethnicity and nationality?	8
Q8	Are descriptions describing from the listeners' perspective?	2
Q9	Are techniques applied in the description?	29
	Total of requirements	75

To use this evaluations, the specialist should verify all accessibility requirements for each alternative content to decide the level, applicability and conformance of the alternative content. Also, the specialist can make comments about the evaluation. Table 10 shows the structure for evaluation composed by a number of the question, accessibility guideline description, level, applicability and conformance. The level is filled with M (Mandatory) or R (Recommended). The applicability is filled with Y (Yes) and N (No). The specialist should verify if the accessibility

Table 9 – Questions for transcript evaluation

Nº	Questions for transcript	Amount of req.
Q1	Does transcript heading include video information?	7
Q2	Are speakers' names included?	5
Q3	Is all speech content included?	3
Q4	Is relevant non-speech audio included?	2
Q5	Is relevant visual information included?	4
Q6	Does transcript have proper punctuation and acronyms definitions?	2
Q7	Does transcript have an adequate edition?	3
Q8	Does transcript allow navigation and clarifications?	2
	Total of requirements	28

requirements is valid to the alternative content that is being evaluated for a specific video. The conformance is filled with Y (Yes) and N (No). If the specialist determine that an accessibility requirement is applied to the video, it is necessary to verify if the accessibility requirement was addressed or not for a specific video. All questions and accessibility requirements for alternative content are described on [Appendix A](#).

Table 10 – Example of requirements for caption

Nro	Guideline	Level	Applicability		Conformance	
		M, R	Y,N	Comments	Y,N	Comments
Q1	Do all text (case, font, line division, caption duration, and caption placement) have a correct appearance and presentation?					
1.1	Mixed case characters are preferred for readability. However, capital letters are used for screaming or shouting.					
1.2	Font characteristics must be consistent throughout the media.					
1.3	The use of white characters is preferred.					

After applying the evaluation, a set of measures should be calculated to obtain a quantitative result. The [Equation 3.1](#) represents the percentage of mandatory requirements covered by the alternative content. The [Equation 3.2](#) represents the percentage of recommended requirements covered by the alternative content.

$$M = \frac{CM}{AM} * 100 \quad (3.1)$$

$$R = \frac{CR}{AR} * 100 \quad (3.2)$$

Where:

M is a measure that represents how much of mandatory requirements are addressed;

R is a measure that represents how much of recommended requirements are addressed;

CM is the total of mandatory requirements addressed (count the number of Y at conformance column);

AM is the quantity of applicable mandatory requirements (count the number of Y at applicability column);

CR is the total of recommended requirements addressed (count the number of Y at conformance column);

AR is the quantity of applicable recommended requirements (count the number of Y at applicability column);

As each alternative content is composed by a set of questions, it is necessary apply the [Equation 3.3](#) and [Equation 3.4](#).

$$AM = \frac{\sum_{i=1}^{i=n} M_i}{n} \quad (3.3)$$

$$AR = \frac{\sum_{i=1}^{i=m} R_i}{m} \quad (3.4)$$

Where:

AM is the average between the mandatory requirements;

AR is the average between the recommended requirements;

n is the quantity of mandatory requirements by each alternative content.

m is the quantity of recommended requirements by each alternative content.

Note that n and m can have different values because some questions can apply only mandatory requirements and none recommended requirements, or may apply only recommended requirements.

To measure how much of accessibility requirements each alternative content addressed to, reference values were developed, as shown in the [Table 11](#). This table is composed by two measures: Mandatory (A, B, or C) and Recommended (+++, ++, or +). The percentages of M and R obtained by QT should be searched on the [Table 11](#) and get a new measure for each alternative content. For example, considering the goal about validation of accessibility requirements for captions, where QT = 78%M and 52%R. According to the [Table 11](#) the final result will be A++, that means captions accomplish most of the mandatory accessibility requirements and more than half of the recommended accessibility requirements.

Table 11 – Quality measure of the accessibility level

Mandatory		
70% <M ≤ 100%	A	Accomplish most of the mandatory accessibility requirements
50% <M ≤ 70%	B	Accomplish more than half of the mandatory accessibility requirements
M ≤ 50%	C	Accomplish some of the mandatory accessibility requirements
Recommended		
70% <R ≤ 100%	+++	Accomplish most of the recommended accessibility requirements
50% <R ≤ 70%	++	Accomplish more than half of the recommended accessibility requirements
R ≤ 50%	+	Accomplish some of the recommended accessibility requirements

3.5 Distribution

At this stage it is important that people with disabilities are considered using an accessible video player. Case studies reported some difficulties that could be improved with the addition of functionalities to the video player. Below these difficulties and their solutions are reported.

- **Low volume.** Older people with auditory impairment reported that low volume is common when they watch movies (see [subsection 3.1.1](#)). *The video player needs to increase volume independently of audio volume control of the computer operating system.*
- **Difficulty with keyboard navigation** (explained in the [subsection 3.1.1](#)). *All functionalities of a video player need to be reachable using keyboard navigation.*
- **Lack of rewind and forward controls.** Users reported the need of reward to some specific point in the video in order to hear the content again (see [subsection 3.1.1](#) and [subsection 3.1.2](#)). A blind user expressed that he misses forward and rewind controls because when he wants to listen again some part of the video, he needs to start from the beginning (see [subsection 3.1.4](#)). *The rewinding and forwarding functionalities are recommended.*
- **Lack of speed down control.** Participants expressed that this functionality can be useful to hear the pronunciation of words and see movements like people dancing or horse races (see [subsection 3.1.2](#)). *Speed down control is recommended.*
- **Lack of search for a scene.** (reported in the [subsection 3.1.1](#)). A participant believes the search control is faster than other controls to search for a specific part of the video (see [subsection 3.1.2](#)). *The video player can provide search of text on captions.*
- **Lack of annotations.** A participant reported she wanted to use annotations on the video because she usually writes the timestamp (see [subsection 3.1.1](#)). A participant said annota-

tions are useful when he wants to return to a specific part of the video (see [subsection 3.1.2](#)). *It indicates that an accessible video player should allow users to create annotations or chapters.*

- **Lack of help control** A participant with auditory disabilities reported she would like other video players to have the help control functionality.
- **Lack of shortcuts.** A participant reported lack of shortcut keys for full screen (see [subsection 3.1.1](#)). *If it is possible, all functionalities of video player should have shortcuts.*
- **The video player is not accessible.** Blind participant reported that he has difficulties to access videos on the Internet because most of video players are developed with Flash technology (see [subsection 3.1.4](#)).

To use the video content on the Web effectively, the video players must provide functionalities according to disabilities:

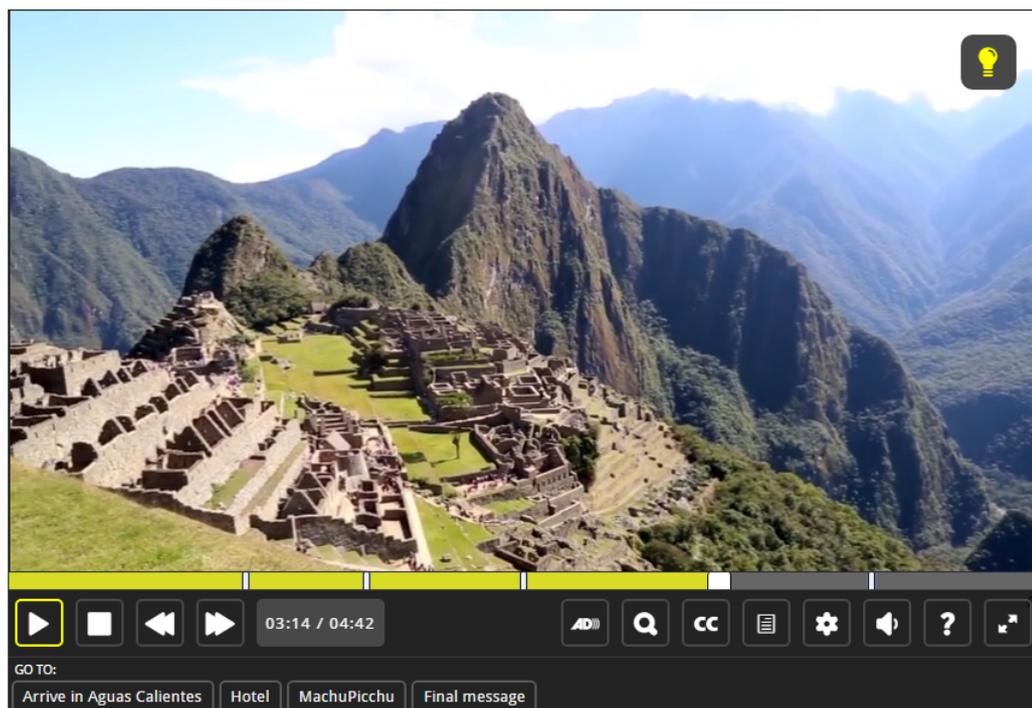
- **Auditory disability.** [Abou-Zahra \(2012\)](#) explains that transcripts and captions of audio content is needed for people with auditory disabilities; also media players should display captions and provide options to adjust the text size and colors of captions, provide options to stop, pause, or adjust the volume of audio content (independently of the system volume). [Carrera et al. \(2013\)](#) suggest the option to remove the ambient sound can be a solution for people with moderate hearing impairments or on situations where the noise is moderate. Also, for some people with auditory disabilities, sign language is their first language, thus, providing information in sign language will help them.
- **Visual disability.** [Abou-Zahra \(2012\)](#) suggests some caption configurations (font size, font color, and spacing) but it can also be important to have the option to change font family and background color of caption. Other important functionalities are increasing and decreasing size of captions and names of buttons, or enabling and disabling audio description. Support to screen reader, keyboard navigation and reading text using refreshable braille provide different ways for interaction.
- **Cognitive and neurological disability.** Enabling and disabling of transcript, closed caption and audio description provide different ways to understand the content. [Abou-Zahra \(2012\)](#) explained that people with cognitive and neurological disabilities use different types of web browsing methods, depending on their particular needs, for example, some people use text-to-speech software to hear the information while reading it visually, or use captions to read the information while hearing it. Other ways of navigation are the search option or links/chapters. Similarly, caption configuration can help to assisting reading and writing.

- **Speech disability.** People with speech disabilities find challenges with voice-based services, such as web applications that are operated using voice commands (ABOU-ZAHRA, 2012). Those people need different ways of navigation, not only based on speech recognition.
- **Physical disability.** Abou-Zahra (2012) explained that people with physical disabilities often use specialized hardware and software such as: ergonomic or specially designed keyboard or mouse; head pointer, mouth stick, and other aids to help typing; on-screen keyboard with trackball, joystick, and switches to operate it; voice recognition, eye tracking, and other approaches for hands-free interaction. It is important that the video player has single keystrokes to activate options.

3.5.1 Facilitas - An accessible video player

We developed an accessible video player, called Facilitas Player, using HTML5, JavaScript, jQuery, jQuery UI and CSS. It uses the HTML5 <video> element for browsers that support them. Facilitas is available on the link <cafe.intermedia.icmc.usp.br:22080/facilitas>. Figure 17 shows the Facilitas interface. Its current controls are: play/pause, stop, rewinding/forwarding of 10 seconds, audio description, search, closed caption, transcript, volume, help, full size, annotations, change in the player language, caption options (font color, font size, font family and background color), and keyboard navigation. Figure 17 shows Facilitas interface.

Figure 17 – Facilitas - An accessible video player



Source: Elaborated by the author.

Table 2 (presented on Chapter 2) shows the functionalities of the video players surveyed. The Facilitas presents most of the functionalities of this table, missing only the caption language and preferences. Following are described the functionalities that have been developed. Most of them are related to guidelines of UAAG 2.0 and ISO 9241-171.

- Play, pause and stop controls (2.11.6 criterion of UAAG and 10.8.2 of ISO 9241-171).
- Volume controller (1.5.1 criterion of UAAG 2.0 and 10.6.2 of ISO 9241-171).
- Viewport resizing or full screen (1.8.3 and 1.8.8 of UAAG 2.0 and 10.5.8 of ISO 9241-171).
- Speed controller (slower and faster).
- Rewind 10 seconds and forward 10 seconds.
- Support closed captions and subtitles (1.1.2 criterion of UAAG 2.0, 10.1.3 and 10.7.2 of ISO 9241-171).
- Support audio description (1.1.2 criterion of UAAG 2.0, 1.2.3 criterion of WCAG 2.0 and 10.1.3 of ISO 9241-171).
- Support transcript.
- Text configuration. Change in font family, font color, and font size (1.4.1 criterion of UAAG 2.0, 1.1.2 criterion of WCAG 2.0 and 10.7.3 of ISO 9241-171). Also the player enables the setting of the background color of the caption.
- Support to keyboard navigation (2.1.1 criterion of UAAG 2.0).
- Emphasis on the selection and focus on the use of the keyboard (1.3.1 and 1.3.2 criterion of UAAG 2.0).
- Search of text on caption (guideline 2.4 of UAAG 2.0). The search control enables the search for a word or phrase that appears in the caption. The player will show all results in a list and when one is selected, the player reproduces the selected part of the video. For instance, if it is searched for the word “popular” and one result was returned, when this result is chosen, the player reproduces the video.
- Annotations. An annotation is a link to some specific time in the video. They are composed of three elements, namely name of the annotation (for example, “Tip: dark chocolate”), description, when the tag is selected (“01:00 - how to make dark chocolate”) and marks in the timeline (tag “Tip: dark chocolate” is the third mark in the timeline). The annotation control enables the developer to add tags to the video, which facilitates the search.

- Light controller. Represented as a lamp icon on the video, enables the player to turn on/off lights. When the light is on, the player is positioned in the middle of the screen and hides the rest of the page content, emphasizing the player. This feature can help people with visual impairment as the brightness outside the player decreases.
- Language of the player. The video player currently has three language Portuguese, Spanish and English.
- Keyboard shortcuts. Located at the Help Control.
- Control description. Each control present a description to facilitate the navigation and memorization.
- Sign language. It was incorporated an online translator from text in Portuguese (transcript) to Libras.

[Source code 1](#) shows the addition of two types of videos, caption, audio description, transcript. [Source code 2](#) shows how to create the annotations, modify the HTML file and list the annotations with the time using 'href', long description in 'title' and short description. [Source code 3](#) shows the instantiation of Facilitas.

Source code 1: Adding the alternative content

```

1
2 <video controls="controls" id="video" width="854" height="480" data-
   taglist="tags" data-audiodescription="video/MachuPicchu_ad.mp3">
3   <track src="video/MachuPicchu_legenda.srt" kind="subtitle"
   srclang="pt-BR" label="Português" />
4   <track src="video/MachuPicchu_transcript.html" kind="transcript"
   srclang="pt-BR" label="Português" />
5   <source src="video/MachuPicchu_480_854.mp4" type="video/mp4" />
6   <source src="video/MachuPicchu_480_854.ogv" type="video/ogg" />
7   <p>Your browser does not support Ogg/Theora.</p>
8 </video>

```

Source code 2: Adding annotations

```

1 <!-- Annotations -->
2 <ul id="tags">
3   <li><a href="#00:09" title="00:09 - Long description">Short
   description </a></li>
4   <li><a href="#00:22" title="00:22 - Another relevant scene">
   Another scene</a></li>
5 </ul>

```

Source code 3: Facilitas instantiation

```
1 $(document).ready(function(){
2   $('video').facilitas({
3     //Principal language:
4     'language'      : 'pt',
5     //Folder of the languages:
6     'baseLangFolder': '/js/'
7   });
8 });
```

[section 2.3](#) presented a literature review about video player research studies, most of them developed accessible video players, and some of them reported tests with real users for specific disabilities. The user interaction with video players helps to identify their difficulties and allow the designer to provide solutions to ease their interaction. Providing alternative content is part of the solution for people with different disabilities, so they can understand video content. However, the most accessed websites that use video players did not provide options to add alternative content, some of them (YouTube and Yahoo) allow the addition of captions. These websites developed the caption functionality in 2014, and in 2015 Facebook allowed the addition of captions. So that the awareness of accessibility is a process that takes some time to be adopted.

3.6 Discussion

This chapter covered a method for alternative content authoring, named Video4All, as part of an accessible video production process. The Video4All method supports amateur authors with the alternative content authoring, through the insertion of activities regarding accessibility recommendations in all stages. Therefore many issues in the alternative content authoring can be avoided considering accessibility from pre-production and production stages. Guidelines for caption, audio description and sign language were found in the literature, but for transcript only recommendations were found. A default template for transcript was developed in order to standardize the transcript file.

The Video4All includes an alternative content evaluation, it was developed based on GQM method and considering accessibility requirements. The evaluator should choice accessibility requirements for an specific video, this choice is vulnerable to subjectivity, therefore it represents a risk. The level (mandatory and recommended) for each accessibility requirements was established by the researcher, a deeper study could be done to improve these levels. Also the percentage to categorize the levels was defined by the researcher, thus, empiric studies are necessary to verify the effectiveness of this proposal.

At distribution stage it was developed an accessible video player named Facilitas. Facilitas provide controls to helping users to find specific content on the video, understanding the video

content through alternative content, changing their preferences about caption format, and others. This player was used in study cases with real users to find difficulties and needs about video player interaction. The lessons learned during the study cases were applied on the activities of the accessible video production process.

To evaluate the Video4All, a case study with authors was conducted. The next chapter explains the alternative content authoring made by amateur authors and the evaluation based on accessibility requirements made by a specialist.

METHOD VALIDATION

This thesis aimed to propose a method for alternative content authoring, called Video4All, described in [chapter 3](#). As Video4All is focused on amateur authors, its validation used a case study method. According to [Yin \(2009\)](#) a case study method should be considered when the focus of the study is a phenomenon in a real context. In this way, a case study with amateur authors was conducted to understand how they create alternative content using the method and without using the method and also to obtain their opinions as a qualitative analysis.

The case study aims to investigate the alternative content authoring (caption, audio description and transcript) by authors with and without following the Video4All method.

This chapter addresses the planning of the case study at [section 4.1](#). The data collection is reported at [section 4.2](#). The analysing and reporting are discussed at [section 4.3](#). Finally, [section 4.4](#) shows a brief discussion about this experimental evaluation.

4.1 Planning

This research observed the Resolution 466/12 of the Ethic Committee on Research Involving Human Subjects at University of São Paulo. The project was approved with *Certificado de Apresentação para Apreciação Ética* (CAAE) nº 45075015.8.0000.0065 and report nº 1.073.606. A copy of the report is available at [Appendix A](#).

In this case study, independent and dependent variables were defined. The independent variables are:

- Method for alternative content authoring (treatment).
- Subject's experience in the generation of alternative content (experimental object).
- Tasks to be performed by subjects (context).

The dependent variables are:

- Amount of accessibility requirements accomplished in each alternative content (caption, audio description and transcript).
- Time to complete, amount of time required for all tasks to be completed.

All subjects were provided with the material described below:

- Informed Consent Form.
- Pre-session questionnaire: form with profile of the subjects, their experience regarding the use of Web systems and knowledge about accessibility for videos.
- Training about video accessibility. Explanation by specialist about concepts of video accessibility and alternative content by means of a video¹. As well as explanation on the basic requirements which a caption, audio description, transcript and sign language need in order to be accessible.
- Training about the method to support the accessible video authoring. It was developed a web site² to explain the method and facilitate the navigation.
- Post-session questionnaire: form for data collection about the interaction to generate alternative content.

Two videos was used during the case study: a short documentary and an animation video. The audio of the videos was in portuguese. The video description are explained on Table 12.

Table 12 – Video's characteristics

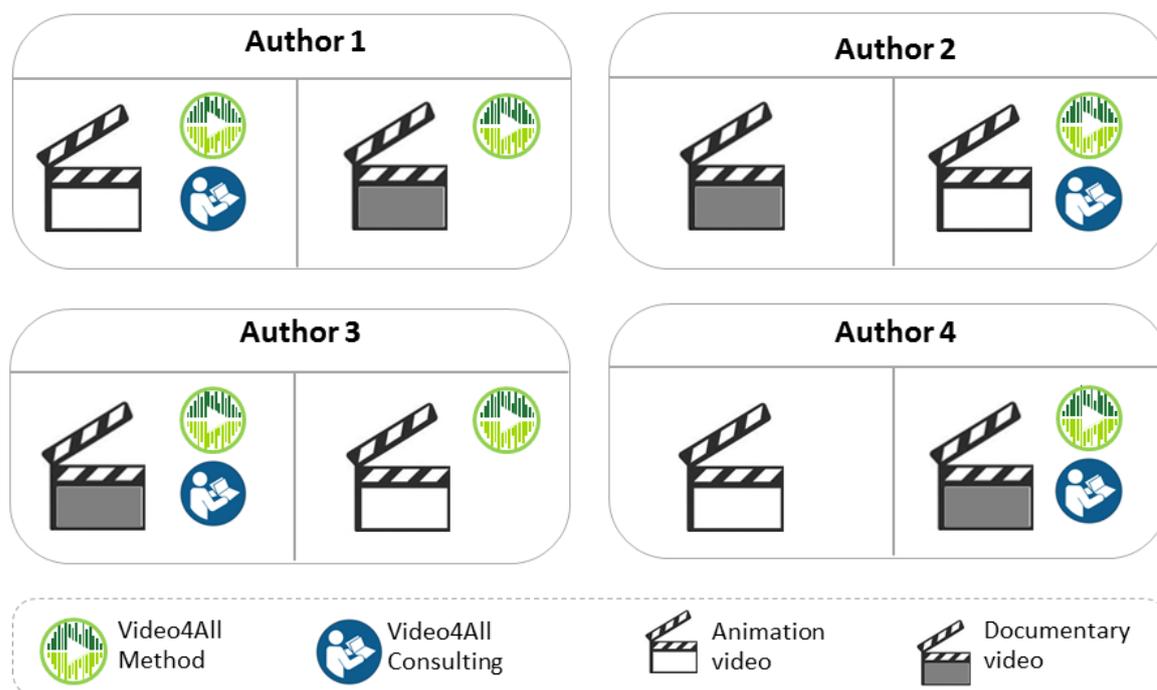
Video	Duration	Abstract
Short documentary	04:42	documentary about Machu Picchu, where two Brazilians share their experience traveling to Machu Picchu
Animation video	05:47	A macaw arrives in <i>Rio de Janeiro</i> and has a conversation with two Brazilians birds

The procedure used 2 (two) videos (a short documentary (grey color in Figure 18) and an animation video (white color in Figure 18). Each participant authored alternative content (caption, audio description and transcript) for both videos. Figure 18 shows the assignment plan for the case study. Authors 1 and 3 learning the Video4All method on the first video and applying the method on the second video. Authors 2 and 4 learning the Video4All method on the second video.

¹ <<https://www.youtube.com/watch?v=KGPdpwtIRjs>>

² <<https://sites.google.com/site/videoaccessivel/>>

Figure 18 – Case Study Planning



Source: Elaborated by the author.

Table 13 – Author's profile

Author	Age	Gender	Experience creating videos	Experience editing videos	Experience creating alternative content
A1	28	F	no	no	no
A2	28	M	yes	yes	no
A3	28	F	yes	yes	no
A4	26	M	yes	yes	no

4.2 Collecting

The researchers made an invitation to ICMC students via an e-mail list. Four (4) students accepted to participate, as content authors, in the case study about the method application. They are Brazilians and all content was in Portuguese. For each author it was individually scheduled a session. Each session was recorded using two applications of Morae software (version 3.3.3), i.e. Recorder and Manager, to facilitate the research process and data analysis. They signed the informed consent form and answered the pre-session questionnaire. Their age ranged from 26 to 28 years old. Two participants were female. They had no disability. Three users (labeled as A1, A2 and A3) never used an accessible video player. Each participant spent approximately two hours on the case study. A1 had never created and edited videos. None of them had created alternative content before. [Table 13](#) shows the author's profile.

All authors were instructed about video accessibility. Following, the procedure used by authors is explained. The significant content generated by the authors was in Portuguese and the

English translation is reliable to the Portuguese content.

- **Author 1.** She learned the method and began the authoring of the first video. She learned the Aegisub tool, wrote some phrases of dialogue, added the character's names to all lines and adjusted the time. She added significant content as "birds playing tambourine", "reading Portuguese dictionary" and "talking with a *Rio de Janeiro* accent". Then, she exported the caption. She used the same significant content to create the audio description phrases. She opened the Soar online software to convert the text to speech. Then opened the GoldWave software, added silences between the phrases, verified if the audio description not overlap the original audio and exported the audio description. She downloaded the transcript template and edited using the Notepad ++ software. She added the video information and copied all text of the caption to the transcript without adding or editing information. She continued the authoring of the second video. She used the same tools learned on the first video and used the same procedures. For caption, she added the significant content "Peruvian music". For audio description, she used the significant content created for caption and created two more: "women sewing" and "woman with Peruvian costumes carrying a child". For transcript, she used the transcript template, added the video information, copied the caption and audio description information without adding or editing information.
- **Author 2.** He used the Notepad software to wrote some phrases of dialogue. He added the character's names without pattern and the initial time for each phrase. He added significant content as "Subway sign Ollantaytambo" and "Andean music", and added other content less or not significant as "People interacting in a market", "Name of the metro company Perurails written in one of the wagons", "detail of a Machu Picchu souvenir - Cusco - Peru", "couple in their hotel", "view from the bedroom window to the city", "couple arrive at the ruins of Machu Picchu and tell about their experience until the arrival on the place", "detail of constructions on various angles". He saved the caption as text file. For audio description, he used the Voice recorder software of Windows. He recorded three phrases using his voice: "Subway sign Ollantaytambo", "Name of the metro company Perurails written in one of the wagons" and "view from the bedroom window to the city". For transcript, he used the caption file and described three significant content with more details, but these content are not concise. Then, he learned the method and continued with the authoring of the second video. He learned the Aegisub tool, wrote some phrases of dialogue, added the character's names and adjusted the time. He added significant content as "yellow bird playing tambourine", "dictionary falls from your hand" and "with a *Rio de Janeiro* accent". Then, he exported the caption. He learned how to use the Soar online software to generate audios from text. He wrote at Soar three significant content: "view from Copacabana beach", "sign of environmental conservation center of Rio de Janeiro" and "Blu is tidy with raised feathers". He opened the GoldWave software, added silences,

verified if the audio description not overlap the original audio and exported the audio description. For transcript, he used the transcript template, added the video information, copied the caption and audio description information, and added description about a scene.

- **Author 3.** She learned the method and began the authoring of the first video. She learned the Aegisub tool, wrote some phrases of dialogue, added the character's names to all lines and adjusted the time. She added significant content as "children buying gelatine", "people at a train station" and "andean music playing". Then, she exported the caption. She used the same significant content to create the audio description phrases. She opened the Soar online software to convert the text to speech. Then opened the GoldWave software, added silences between the phrases, verified if the audio description not overlap the original audio and exported the audio description. She downloaded the transcript template and edited using the Notepad ++ software. She added the video information and copied all text of the caption to the transcript, and added description about a scene. She continued the authoring of the second video. She used the same tools learned on the first video and used the same procedures. For caption, she added the significant content "birds singing and playing", "with a *Rio de Janeiro* accent", "Blue drops the dictionary", "car walking along the seashore avenue", and "sign: environmental conservation center - Rio de Janeiro". For audio description, she used three significant content created for caption: "birds singing and playing", "car walking along the coastal avenue", and "sign: environmental conservation center - Rio de Janeiro". For transcript, she used the transcript template, added the video information, copied the caption and audio description information, and added description about a scene.
- **Author 4.** He used the Notepad software to wrote some phrases of dialogue. He added the character's names for each phrase. He added significant content as "singing" and "talking with a *Rio de Janeiro* accent". He saved the caption as text file. For audio description, he used the Voice recorder software of Windows. He recorded three phrases using his voice: "playing tambourine and dancing on the roof of a yellow car", "leafing through Portuguese book" and "the yellow car start from traffic light and transiting along the coastal avenue in Rio". For transcript, he used the caption file, added significant content and described some phrases with more details. Then, he learned the method and continued with the authoring of the second video. Different from the other authors, the author 4 created first the transcript, then the caption and finally the audio description. For transcript, he used the transcript template, added the video information, wrote some phrases of dialogue, added the character's names and description of some scenes. He learned the Aegisub tool, used the transcript text and adjusted the time. He added significant content as "Typical music playing", "staggering" and "woman laughing". Then, he exported the caption. He learned how to use the Soar online software to generate audios from text. He generated three audios: "on a sidewalk, children buy juices", "at market, people buy fruits" and

Table 14 – Accessibility requirements evaluation for caption

Caption		
Questions	Requirements applied	Requirements accomplished
Q1	1.1 (R), 1.8 (M), 1.9 (M), 1.10 (M)	1.1, 1.8, 1.9, 1.10
Q2	2.1 (M), 2.2 (M), 2.3 (M), 2.6 (M), 2.9 (M), 2.13 (M)	2.1, 2.2, 2.3, 2.9, 2.13
Q3	3.1 (M)	3.1
Q4	4.1 (M), 4.2 (R), 4.7 (M), 4.10 (M), 4.11 (M)	4.1, 4.2, 4.10, 4.11
Q5	5.2 (M), 5.3 (M), 5.4 (M), 5.5 (M)	5.2, 5.3
Q6	6.1 (M), 6.2 (M)	6.1, 6.2
Q7	7.2 (M)	-
Q8	8.3 (M)	-
Q9	-	-

“at Ollantaytambo station, people are waiting to get into a blue train naming Perurail”. He opened the GoldWave software, added silences, verified if the audio description not overlap the original audio and exported the audio description.

4.3 Analysing and reporting

When authors learned the method on the first video and applied it on the second video, they spent 46 minutes on average for both videos. While they did not use any method on the first video and used our method on the second video, they spent 67 minutes on average for both videos.

To make the accounting of accessibility requirements accomplished for each alternative content, the analysis employed the alternative content evaluation explained on [item 3.4.3](#) on the previous chapter. This evaluation was made by a specialist using the alternative content produced by each author.

For each alternative content, the specialist analyzed which requirements were applied according the video fragment that each author chose and which requirements accomplished. The requirements applied for the same video can be different because the author chose different fragment for generate the caption, audio description or transcript. An example of this exception can be seen on the caption of the animation video, the requirements applied by the author 1 on video 1 are different those applied by the author 3 on video 2. [Table 14](#) shows the accessibility requirements evaluation for caption generated by author 1 on the first video. The complete evaluation is presented in [Appendix B](#).

After applying the evaluation, it was applied the equations [3.1](#) and [3.2](#) for each question. For example, considering the data of [Table 14](#) on Question 4, the results obtained are explained

Table 15 – Percentages of mandatory and recommended requirements covered by caption

Caption		
Questions	%M	%R
Q1	100	100
Q2	83,3	-
Q3	100	-
Q4	75	100
Q5	50	-
Q6	100	-
Q7	0	-
Q8	0	-
Q9	-	-

on equations 4.1 and 4.2.

$$M = \frac{CM}{AM} * 100 = \frac{3}{4} * 100 = 75\% \quad (4.1)$$

$$R = \frac{CR}{AR} * 100 = \frac{1}{1} * 100 = 100\% \quad (4.2)$$

After applying the equations 3.1 and 3.2 for each question, it was obtained the Table 15 with all percentages of mandatory and recommended requirements covered by caption.

Then, it was applied the equations 3.3 and 3.4 for each alternative content. For example, equations 4.3 and 4.4 show the average between the mandatory and recommended requirements.

$$AM = \frac{\sum_{i=1}^{i=n} M_i}{n} = \frac{\sum_{i=1}^{i=8} M_i}{8} = 64\% \quad (4.3)$$

$$AR = \frac{\sum_{i=1}^{i=2} R_i}{2} = 100\% \quad (4.4)$$

Table 16 shows the percentages of accessibility requirements for each alternative content considering the two videos (V1 and V2) and four authors (A1, A2, A3, and A4). For example, when author 1 (A1) was working with the video 1 (V1) using the method, A1V1 obtained 64% of the mandatory (M) accessibility requirements (obtained on Equation 4.3) and 100% of the recommended (R) accessibility requirements (obtained on Equation 4.4) for caption, 63% of M and 88% of R for audio description, 83% of M and 67% of R for transcript, and the arithmetic mean was 70% of M and 85% of R of the accessibility requirements.

On Table 16, authors 1 and 3 learned the method on the first video and applied it on the second video without consulting the method. While authors 2 and 4 generated alternative content without the method on the first video and learned the method on the second video. The results show that the percentages of mandatory and recommended requirements on the second video are

Table 16 – Accessibility requirements achieved by authors

Author	Caption		Audio description		Transcript		Arithmetic mean	
	%M	%R	%M	%R	%M	%R	%M	%R
A1 V1	64%	100%	63%	88%	83%	67%	70%	85%
A1 V2	79%	100%	88%	93%	92%	67%	86%	87%
A2 V1	55%	50%	81%	43%	58%	67%	65%	53%
A2 V2	97%	100%	93%	88%	92%	67%	94%	85%
A3 V1	90%	100%	87%	100%	100%	67%	92%	89%
A3 V2	92%	100%	88%	100%	100%	100%	93%	100%
A4 V1	83%	100%	86%	50%	83%	67%	84%	72%
A4 V2	90%	100%	87%	73%	100%	100%	92%	91%

equal or greater than the percentages on the first video. In other words, the authors have gained experience and applied the concepts learned in the first video.

To understand the accessibility quality for each alternative content generated by authors, an analysis was performed using a rating scale, as reported in Table 11 from the previous chapter in order to elaborate Table 17. This type of measure helped the author to understand the accessibility level of quality accomplished by each alternative content. For example, when author 1 was working with video 1 using the method, A1V1 obtained B+++ for caption, that means captions accomplish more than half of the mandatory accessibility requirements and most of the recommended accessibility requirements.

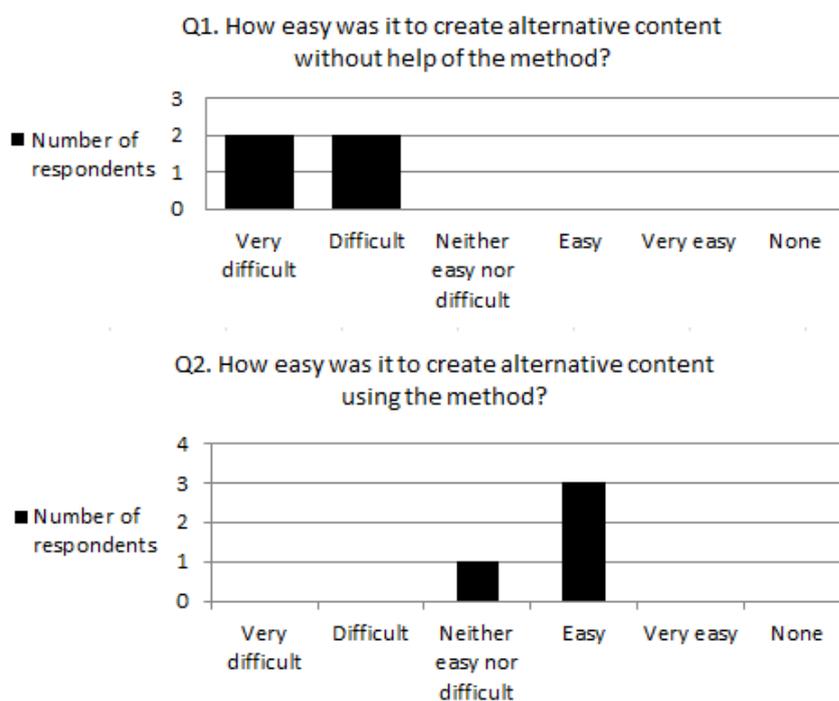
Table 17 – Results of the accessibility quality for each alternative content generated by authors

Author	Caption	Audio description	Transcript
A1 V1	B+++	B+++	A++
A1 V2	A+++	A+++	A++
A2 V1	B++	A+	B++
A2 V2	A+++	A+++	A++
A3 V1	A+++	A+++	A+++
A3 V2	A+++	A+++	A+++
A4 V1	A+++	A++	A++
A4 V2	A+++	A+++	A+++

The utility of the Video4All method has also been reflected on the answers to the questionnaire (see Figure 19). When asked how easy it was to create alternative content without help of the method (Q1), A1 and A2 said it was very difficult and A3 and A4 said it was difficult. Author 2 said that “*finding the exact time to do the insertion was difficult without the method*”. It was because when the participants create alternative content without the method, they usually do not know any tool to edit the caption, they use the notepad. Author 2 only learned the method in the second video. When they were asked how easy it was to create alternative content using the method (Q2), the respondents said it was easy (3 of them) and neither easy nor difficult (1 of them).

Author 3 expressed such appreciation by saying that “*I had no previous experience about the subject, so I found it practical and easy to learn the method and manipulate the tools*”. Author 3 had previous experience creating and editing videos, but she had never created alternative content. Also on the first video she learned the method and tools.

Figure 19 – Results from the alternative content activity using the proposed method



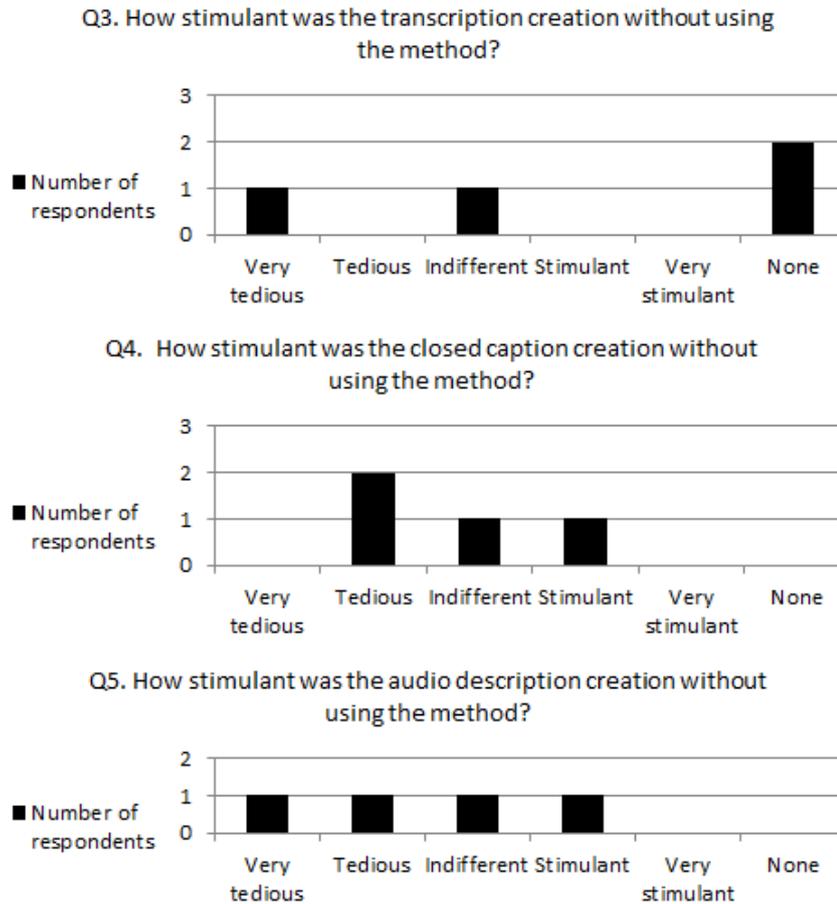
Source: Elaborated by the author.

Some participants reported difficulties regarding the development of alternative content. Author 3 mentioned that “*the process is a bit laborious because you have to access programs with different interfaces, but I believe that making an integration would make it easier*”. Author 1 said “*the problem was the duplication of data when creating files separately and after that, integrating them in another external tool*”. Authors 1 and 3 made the authoring using the method on the first video, therefore on the second video they had learned tips from the method and applied them. Authors 2 and 4 did not complain about the laborious maybe because they observed that made the authoring without using the method is worst. However, Author 4 reported “*it was difficult to made a transcript of colloquial language, slang and expressions*”. He found some difficulties to understand the conversation between two actors because they were talking at the same time.

A similar feedback was obtained on the questions regarding participants’ impression about each alternative content creation without using the method (see Figure 20) and using the method (see Figure 21). When asked how stimulant the transcript/closed caption/audio description creation was without using the method, most respondents said it was very tedious, tedious or indifferent (Q3, Q4 and Q5).

Regarding developing captions, Authors 2 and 4 answered it was tedious. It is important

Figure 20 – Results from the alternative content creation without using the method



Source: Elaborated by the author.

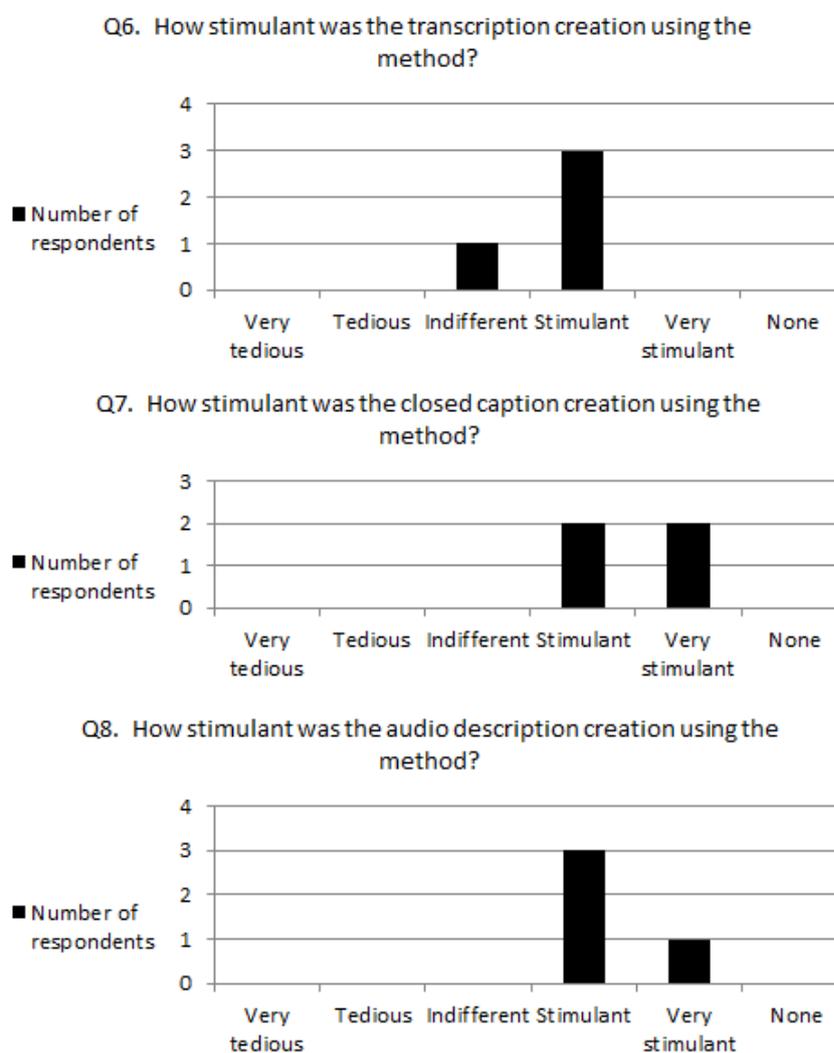
to highlight that they developed the captions without the method on the first video, while authors that learned the method on the first video (Authors 1 and 3) answered stimulant and indifferent.

Regarding developing transcript, Author 2 answered it was very tedious and Author 4 indifferent. Authors 1 and 3 answered none because they said they learned the method on the first video and they did not know without the method. This affirmation suggests that because authors have learned the method on the first video, they applied it on the second video.

Developing audio description was a difficult task for Author 1, tedious for Author 2, indifferent for Author 3 and stimulant for Author 4. Author 1 was the only author that never created or edited videos, she learned the method on the first video, and she had difficulties to use the audio editing tool. When asked how stimulant the transcript/closed caption/audio description creation was using the method, most respondents said it was stimulant or very stimulant (Q6, Q7 and Q8). Author 1 expressed that *“I like it because allows the video creation to people with disabilities, videos like that are not commonly found”*. She was very interested in video accessibility theme, she did not had knowledge about that.

All participants said closed caption was easy to create, using the method. Author 1

Figure 21 – Results from the alternative content creation using the method



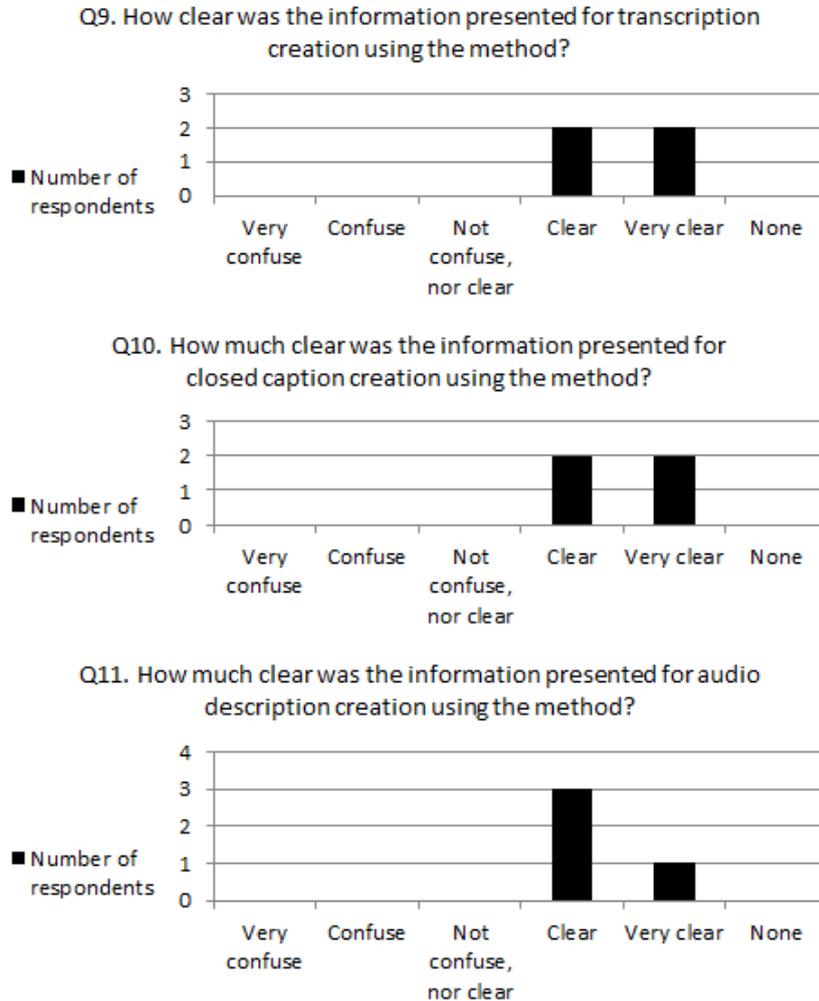
Source: Elaborated by the author.

reported “*the tool was easy to use*”, Author 2 commented “*it was easy to put information correctly*”, Author 3 said “*the guidelines are well explained*” and Author 4 mentioned “*the films that I have watched helped me with the caption creation*”. However, Author 2 reported “*it is complicated to know the start and the end of an insertion*”. Although they said the process to create captions was easy, they spent time understanding the guidelines, searching the specific guideline appropriate for the video content, and learning how to use the tool.

None of the participants had made a transcript before and they were not provided with a tool to create transcript. Author 1 said that “*transcript allows adding more information and customize the information*”, while Author 3 commented “*easy process to perform, perhaps could improve the interface. I suggest the creation of an interface to edit the HTML, because the interface would be friendlier to people who did not know to edit this type of document*”. Authors 1 and 3 used the method on the first video, thus, they were provided with a transcript template (explained at [section 3.4.3](#) and illustrated at [Appendix C](#)) in HTML format. Because of that,

Author 3 suggests an interface to edit the transcript.

Figure 22 – Author’s opinion about the information presented in the method



Source: Elaborated by the author.

For some participants the audio description creation was easy using the method. Author 3 commented “*easy to generate and to incorporate to the tool*” and Author 2 said “*efficiency at capturing and correcting insertion*”. Author 3 talked about the automatic speech generation from text, provided by the online tool Soar. Before inserted the text into the tool, she created it searching by the specific guidelines according to the video content. Author 2 commented about the efficiency using the text-to-speech tool and correcting insertion using the audio editing tool. As Author 2 created the AD without using the method on the first video, he explained that using the method was more efficient and provides a correct insertion. For Author 1 the AD creation was not easy “*it is not easy, it is difficult*”, and Author 4 mentioned “*describing the content in appropriate time and succinctly was difficult, as well as separating relevant elements*”. Author 1 reported that using audio editing tool was difficult, matching with the answer of Q5. While Author 4 explained that the process to identify which are the relevant information to describe was difficult.

In general, authors' feedback regarding the Video4All method was positive. When asked how clear the information presented was for each alternative content creation using the method (see [Figure 22](#)), all participants said it was very clear or clear (Q9, Q10 and Q11). Three of the authors (1, 2 and 3) had no previous knowledge about video accessibility, they needed to learn about it and also to create alternative content based on the presented information. Although authors said that the presented information was clear, they complained that learning guidelines, with a lot of information, results a laborious task and they spent time reading it to do the authoring. Some of them suggest to use pictures to explain the guidelines.

In fact, the generation of alternative content requires to spend time for learning the guidelines, but when they are applied on the second video, authors spent less time to generate the alternative content.

Observing the author interactions, most of the authors copied the caption and audio description information to compose the transcript. This suggested the creation of a tool to help with this process. In fact, based on that we have developed the Transcript4All tool (described on [Appendix D](#)).

4.4 Discussion

This chapter described the evaluation of the Video4All method through a case study with four authors. The case study was designed as an scientific experimental procedure, and the stages of such experiment were well defined and accomplished.

In the case study, the authors created three alternative content (caption, audio description and transcript) with and without using the method. The results show that when using the method, more accessibility requirements are accomplished by each alternative content.

The case study with authors provided information about how authors generate alternative content by using or not, the method. The goal of this study was mainly focused on observing the efforts by the authors, during the experimental proposed activity, regarding to accessibility issues that should be addressed at the final alternative content of the video. The developed alternative content, then, should be measured in terms of accessibility requirements. The basis for this observation was specified in the [chapter 3](#), according to the mandatory (M) and recommended (R) requirements that were developed by the authors. It's worth to notice that the impressions and comments given by the authors could reinforce the utility of the Video4All method.

Concerning to the collected data from the case study, the amount of accessibility requirements accomplished on the second video, after learning the method on the first video, is greater than or equal to the amount of accessibility requirements accomplished on the first video without using the method. Furthermore, the accessibility level of a video using the method on the first video is less than or equal to the accessibility level of a video using the method on the second

video.

Most of authors reported that by using the method, the tasks related to generate alternative content on videos were easier than without using it. Although they commented that learning guidelines demanded time to start with the alternative content generation, they recognized that it is a natural process.

Thanks to the case study, we have also designed and developed a transcript tool (Transcript4All) to help users with the transcript generation. Some authors realized the rework when they created the transcript, and they suggested an interface to help with the transcript creation.

CONCLUDING REMARKS

The growth of the available videos on the Internet over the last years is evidence that videos play an important role in society. However, most videos do not show alternative content or even most video players do not provide accessible features. These problems could be due to the lack of knowledge of amateur authors about how to provide accessibility to a video.

Providing accessibility to videos for a heterogeneous audience does not only mean producing videos following guidelines, but also offering resources for amateur authors, so they can overcome barriers to produce accessible videos. Therefore, accessibility should be considered from the beginning of the video production process, since the pre-production stage to prevent issues in the future stages, until distribution stage to show the video in an accessible video player with the desired features. Including accessibility into the video production process does not only help amateur authors, but also benefits end users and promotes a more inclusive society.

This thesis shares the idea that videos are more and more present in people's lives and they will need to comply with the needs of more users, regardless of their characteristics or disabilities. Thus, this thesis developed a method for alternative content authoring as part of an accessible video production process. Considering accessibility from the beginning of the video production process helps with the alternative content incorporation. For example, fast captions can be prevented in the scripting, or the addition of enough silence for audio description also in the scripting.

The Video4All method was an initiative to support the alternative content authoring following the alternative content guidelines and also establishing a way to evaluate this content.

This chapter first answers the thesis questions, then list the main contributions of the thesis, explain the future work and finally list the results' dissemination.

5.1 Answering the thesis questions

Following are discussed the questions initially established.

1. Which elements are necessary for a process to support accessible video authoring capable of promote accessibility? Basically, three elements are necessary: the media (video) should be accessible by itself, alternative content should be included, and an accessible video player should be used. The three first stages of accessible video production process promote accessibility into the media, providing activities to prevent future accessibility barriers, editing, and validation of the video against seizure risks. The Video4All method provide activities for alternative content authoring and evaluation. And tests with users (see [section 3.1](#) and [subsection 3.5.1](#)) showed that the use of an accessible video player can overcome barriers for heterogeneous people.

2. How can a method increase the accessibility level of a video?

As observed on the case study with authors, the videos in which the method was used accomplish more accessibility requirements than the videos where it was not used. When the authors used the method on the first video, they used the same method on the second one, increasing the quantity of accessibility requirements and reducing the time of completion. When the authors on the first video created the alternative content without using the method, there was a significant increase of the quantity of accessibility requirements on the second video.

3. How can a method be included as part of the accessible video authoring process?

The Video4All method is located in the pos-production stage, but in all previous stages the method considered activities that could help with the addition of alternative content. Also, the method helps not only with the production of alternative content following activities, but also with the evaluation. This evaluation presents accessibility requirements to measure the final alternative content.

4. How to evaluate whether the levels of accessibility of the alternative content are compliant with the guidelines established?

To evaluate the accessibility requirements that accomplish each alternative content, the method proposed a categorization that provides a quality measure of the accessibility level. This evaluation helps authors to understand which accessibility level has associated alternative content.

Finally, according to the main research question, formulated to this thesis, the above answers states that: *the accessibility guidelines and an accessible video production process that include them can be combined in a method (Video4All) and this method generates a video with better accessibility levels as result.*

In fact, it was observed in [chapter 4](#) the accessibility level of a video using the method Video4All is higher than the accessibility level of a video without using it. Furthermore, the

accessibility level of a video using the method on the first video is less than or equal to the accessibility level of a video using the method on the second video.

5.2 Main contributions

The most expressive contribution of this work is the Video4All method, presented in [subsection 3.4.3](#) of [chapter 3](#). This method was defined to support the alternative content authoring taking into consideration guidelines and activities for the authoring of each alternative content. Also, the method proposes a quality measure of the alternative content as a way of evaluation. The Video4All method by itself cannot ensure video accessibility, the method needs to be inserted into an accessible video production process. Thus, recommendations about accessibility issues are explained in all stages of the process.

The Video4All method was validated using the case study methodology with real users on [chapter 4](#). This validation allowed understanding that amateur authors can create alternative content based on activities. The alternative content evaluation explained on [item 3.4.3](#) of [chapter 3](#) was applied to the case study, and each alternative content created by authors was evaluated (see [Appendix B](#)).

Additionally, an accessible video player named Facilitas was developed (see [subsection 3.5.1](#)), that was used in tests with users to collect barriers that users have during the interaction with video players and also barriers with the own video product (see [section 3.1](#)). During the accessible video production process, issues observed during the users' interaction with Facilitas were incorporated.

Another contribution arisen of this thesis was an authoring tool named Transcript4All. This tool provides the creation of transcript from caption file or without it, recommendations for creation of transcript, functionalities as add, join, delete, etc. phrases, and file exportation. It was designed from the needs reported by users of the case study explained on [chapter 4](#). More details of the Transcript4All tool are at [Appendix D](#).

5.3 Limitations and future work

With the development of this research, it can be seen that despite positive evidence obtained through the evaluations, there is a lot of study and development to be done in research about accessibility and videos. In this context, some limitations and future work can be highlighted.

One limitation that can make the use of the Video4All method difficult is that author has the necessity to learn all guidelines and activities, requiring a high level of concentration. As observed on case studies, some authors declared that they would need to read the guidelines again

for future alternative content authoring. In this research the focus was on providing all activities necessary to create an appropriate alternative content without worrying about the learning.

In the accessible video production process described in this thesis, the main goal was describing activities that incorporate accessibility to prevent future mistakes that could be considered in early stages. A limitation of the process is that it does not described more in depth the production stage, and also the research proposed future tests with users using all stages of the process. In this thesis, the Video4All method was tested.

During the process, all recommendations and activities are related to prerecorded videos, live videos were not considered. Also, the main author of the thesis was amateur, it was not considered professional authors.

All alternative content (caption, audio description, transcript and sing language) can be described using a declarative language as SMIL or NCL. The Video4All method did not consider the automatic junction of this alternative content to reproduce with the video, the author should choose manually the files and adequate to reproduce in an accessible video player.

The thesis proposed a tool that summarizes the Video4All method, from the project design, video information, addition of caption, audio description, transcript, sing language and annotations. There are tools that help with the production of captioning and transcript, but still they are not completed and neither reuses the information of caption or audio description to incorporate in the transcript.

The choice of accessibility requirements for a specific video by evaluators is vulnerable to subjectivity, therefore it represents a risk. To avoid such risk the method recommends automatization of the calculation of the metrics. This tool standardizes the choices made by the evaluators selecting the specific guidelines related to the video content. For example, if the video presents a sequence with music, the guidelines related to music should be selected and the evaluator will use these guidelines to apply the evaluation.

Another tool that would be useful so the author will not forget how to apply any guideline is developing a tool with an interface that provides options with guidelines implemented. For example the author selects the option insert music, and the syntax is provided for this type of content.

The level (Mandatory and Recommended) for each accessibility requirements in [Table 10](#) was established by the researcher. A deeper study with users could be done, so that final users can weigh up the accessibility requirement, to know which are more important. Based on that, it could be redefined the mandatory and recommended level.

The percentage to separate the levels was defined by the researcher in [Table 11](#) (quality measure of accessibility level for alternative content). More empiric studies to define the percentage of these levels are necessary to verify the effectiveness of this proposal.

Other limitation was the number of participants involved in the case study. The case study was composed by four authors. With more number of participants, quantitative results would be obtained and generalized.

Finally, a future work that can be useful for amateur authors is the development of an automatized tool to support all video production process considering accessibility.

Videos will be more popular in the next years, increasing the videos advertising by Internet and the population became familiarized with them. Nowadays most of the videos are not accessible for all type of audience, but it is a culture that is going to increase. The case study with authors provide evidence of the interest that they have to make an accessible content. With more resources and tools these people feel motivated to produce an accessible video.

5.4 Results' Dissemination

The results of this research work have been published in different conferences.

- Rosas-Villena, J.; Ramos, B.; Goularte, R.; Fortes, R. Video accessibility on the most accessed websites - a case study regarding visual disabilities. In: ANTONA, M.; STEPHANIDIS, C. (Ed.). *Universal Access in Human-Computer Interaction*, 17th International Conference on Human-Computer Interaction (HCII). *Access to Today's Technologies*. Springer International Publishing, 2015, (Lecture Notes in Computer Science, v. 9175). p. 231–241. ISBN 978-3-319-20677-6.
- Facilitas Player wins the 2nd prize at the All@Web Program. *Web Projects: Applications and assistive technologies of All@Web subcategory*. All@Web Web Accessibility Recognition Program is an initiative of W3C Brazil Office. 2014.
- Villena, J. M. R.; Ramos, B. C.; Fortes, R. P.; Goularte, R. An accessible video player for older people: Issues from a user test. *Procedia Computer Science* , v. 27, n. 0, p. 168 – 175, 2014. ISSN 1877-0509. 5th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-exclusion, (DSAI) 2013.
- Villena, J. M. R.; Ramos, B. C.; Fortes, R. P. M.; Goularte, R. Web videos – concerns about accessibility based on user centered design. *Procedia Computer Science* , v. 27, n. 0, p. 481 – 490, 2014. ISSN 1877-0509. 5th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-exclusion, (DSAI) 2013.
- Villena, J. M. R.; Goularte, R.; Fortes, R. P. de M. A user test with accessible video player looking for user experience. In: STEPHANIDIS, C.; ANTONA, M. (Ed.). *Universal Access in Human-Computer Interaction*, 16th International Conference on Human-Computer Interaction (HCII) . *Design for All and Accessibility Practice*. Springer International

Publishing, 2014, (Lecture Notes in Computer Science, v. 8516). p. 623–633. ISBN978-3-319-07508-2.

- Rosas-Villena, J. M. ; Ramos, B. C. ; Fortes, R. P. M. ; Goularte, R. . Facilitas: an accessible video player. In: 19th Brazilian Symposium on Multimedia and the Web (WebMedia), 2013, Salvador (Bahia) - Brazil. XII Workshop de Ferramentas e Aplicações (WFA), 2013. v. I. p. 49-52.

BIBLIOGRAPHY

ABLEPLAYER. **Able Player**. 2015. Accessed: 2015-07-08. Available: <<http://ableplayer.github.io/ableplayer/>>. Cited 2 times on pages 25 and 44.

ABNT. **NORMA BRASILEIRA ABNT NBR 15290 - Acessibilidade em comunicação na televisão**. 2005. Available: <<http://www.crea-sc.org.br/portal/arquivosSGC/NBR2015290.pdf>>. Accessed: 29/06/2015. Cited on page 94.

ABOU-ZAHRA, S. **How People with Disabilities Use the Web**. 2012. Available: <<http://www.w3.org/WAI/intro/people-use-web/diversity>>. Accessed: 29/06/2015. Cited 6 times on pages 35, 72, 78, 82, 100, and 101.

ADC. **Standards for Audio Description and Code of Professional Conduct for Describers**. 2009. Available: <http://www.audiodescriptioncoalition.org/adc_standards_090615.pdf>. Accessed: 09/11/2015. Cited on page 203.

AFB-ACCESSIBLEPLAYER. **Download the AFB Accessible Player: The American Foundation for the Blind's Accessible HTML5 Video Player (Beta)**. 2015. Available: <<http://www.afb.org/info/programs-and-services/technology-evaluation/creating-accessible-websites/download-afbs-accessible-html5-video-player/1235>>. Accessed: 2015-07-08. Cited 2 times on pages 25 and 44.

AMORIM, G. F.; SANTOS, J. A. F. dos; MUCHALUAT-SAADE, D. C. Adaptive layouts and nesting templates for hypermedia composite templates. In: **Proceedings of the 21st Brazilian Symposium on Multimedia and the Web**. New York, NY, USA: ACM, 2015. (WebMedia '15), p. 189–196. ISBN 978-1-4503-3959-9. Available: <<http://doi.acm.org/10.1145/2820426.2820442>>. Cited on page 43.

ANTHONY, L.; KIM, Y.; FINDLATER, L. Analyzing user-generated youtube videos to understand touchscreen use by people with motor impairments. In: **Proceedings of the SIGCHI Conference on Human Factors in Computing Systems**. New York, NY, USA: ACM, 2013. (CHI '13), p. 1223–1232. ISBN 978-1-4503-1899-0. Available: <<http://doi.acm.org/10.1145/2470654.2466158>>. Cited on page 44.

ANTUNES, D.; GUIMARÃES, C.; GARCIA, L.; OLIVEIRA, L.; FERNANDES, S. A framework to support development of sign language human-computer interaction: Building tools for effective information access and inclusion of the deaf. In: **Research Challenges in Information Science (RCIS), 2011 Fifth International Conference on**. [S.l.: s.n.], 2011. p. 1–12. ISSN 2151-1349. Cited 2 times on pages 25 and 43.

ARAÚJO, T. M. U. D.; FERREIRA, F. L. S.; SILVA, D. A. N. S.; OLIVEIRA, L. D.; aO, E. L. F.; DOMINGUES, L. A.; MARTINS, V. F.; PORTELA, I. A. C.; NÓBREGA, Y. S.; LIMA, H. R. G.; FILHO, G. L. S.; TAVARES, T. A.; DUARTE, A. N. An approach to generate and embed sign language video tracks into multimedia contents. **Inf. Sci.**, Elsevier Science Inc., New York, NY, USA, v. 281, p. 762–780, Oct. 2014. ISSN 0020-0255. Available: <<http://dx.doi.org/10.1016/j.ins.2014.04.008>>. Cited 2 times on pages 25 and 43.

BARBOSA, S.; SILVA, B. **Interação Humano-Computador**. Brazil: Campus-Elsevier, 2010. ISBN 978-85-352-3418-3. Cited on page 78.

BASILI, V.; WEISS, D. A methodology for collecting valid software engineering data. **Software Engineering, IEEE Transactions on**, SE-10, n. 6, p. 728–738, Nov 1984. ISSN 0098-5589. Cited on page 95.

BERGER, J. **Ways of Seeing**. Penguin Adult, 2008. (Modern Classics). ISBN 9780141035796. Available: <<https://books.google.com.br/books?id=ilu4K-Xp0v4C>>. Cited on page 23.

BLENDER. **Sintel**. 2010. Available: <<https://durian.blender.org/>>. Accessed: 2014-12-03. Cited 2 times on pages 66 and 76.

BRAULT, M. W. **Americans With Disabilities: 2010**. 2012. Available: <<http://www.census.gov/prod/2012pubs/p70-131.pdf>>. Accessed: 12/02/2016. Cited on page 24.

BULTERMAN, D. C. A.; HARDMAN, L. Structured multimedia authoring. **ACM Trans. Multimedia Comput. Commun. Appl.**, ACM, New York, NY, USA, v. 1, n. 1, p. 89–109, Feb. 2005. ISSN 1551-6857. Available: <<http://doi.acm.org/10.1145/1047936.1047943>>. Cited on page 43.

CAB. **Closed Captioning Standards and Protocol for Canadian English Language Television Programming Services**. 2012. Available: <http://www.cab-acr.ca/english/social/captioning/cc_standards.pdf>. Accessed: 29/06/2015. Cited 3 times on pages 86, 142, and 144.

CARRERA, X.; COIDURAS, J.; SALSE, M.; RIBERA, M.; GRANOLLERS, T. **Guía de contenido digital accesible: vídeo**. España: Universitat de Barcelona and Universitat de Lleida, 2013. ISBN 978-84-695-6918-4. Cited 11 times on pages 25, 37, 38, 40, 43, 46, 51, 77, 78, 81, and 100.

CEN/CENELEC/ETSI. **The European Directive Accessibility requirements suitable for public procurement of ICT products and services in Europe**. 2014. Available: <http://www.etsi.org/deliver/etsi_en/301500_301599/301549/01.01.01_60/en_301549v010101p.pdf>. Accessed: 22/06/2015. Cited on page 30.

CENTER, P. R. **Photo and Video Sharing Grow Online**. 2013. Available: <<http://www.pewinternet.org/2013/10/28/photo-and-video-sharing-grow-online/>>. Accessed: 15/10/2015. Cited on page 24.

_____. **News video on the web**. 2014. Available: <<http://www.journalism.org/2014/03/26/news-videoon-the-web/>>. Accessed: 15/07/2015. Cited on page 24.

CENTER'S, T. **Photosensitive Epilepsy Analysis Tool (PEAT)**. 2015. Available: <<http://trace.wisc.edu/peat/>>. Accessed: 29/06/2015. Cited on page 83.

CHAMPIN, P.-A.; ENCELLE, B.; EVANS, N. W. D.; O-BELDAME, M.; PRIÉ, Y.; TRONCY, R. Towards collaborative annotation for video accessibility. In: **Proceedings of the 2010 International Cross Disciplinary Conference on Web Accessibility (W4A)**. New York, NY, USA: ACM, 2010. (W4A '10), p. 17:1–17:4. ISBN 978-1-4503-0045-2. Available: <<http://doi.acm.org/10.1145/1805986.1806010>>. Cited 2 times on pages 25 and 43.

CHENG, K.-Y.; LUO, S.-J.; CHEN, B.-Y.; CHU, H.-H. Smartplayer: User-centric video fast-forwarding. In: **Proceedings of the SIGCHI Conference on Human Factors in Computing Systems**. New York, NY, USA: ACM, 2009. (CHI '09), p. 789–798. ISBN 978-1-60558-246-7. Available: <<http://doi.acm.org/10.1145/1518701.1518823>>. Cited 2 times on pages 25 and 44.

CISCO. **Cisco Visual Networking Index: Forecast and Methodology, 2014-2019 White Paper**. 2014. Available: <http://www.cisco.com/c/en/us/solutions/collateral/service-provider/ip-ngn-ip-next-generation-network/white_paper_c11-481360.html>. Accessed: 11/02/2016. Cited on page 23.

COMMISSION, F. C. **21st Century Communications and Video Accessibility Act (CVAA)**. 2010. Accessed: 2014-12-05. Available: <<http://www.fcc.gov/guides/21st-century-communications-and-video-accessibility-act-2010>>. Cited on page 32.

COMPESI, R.; GOMEZ, J. **Introduction to Video Production: Studio, Field, and Beyond**. United States: Focal Press, 2005. ISBN 978-0205361076. Cited 5 times on pages 25, 37, 39, 41, and 42.

DCMP. **Captioning key - Guidelines and Preferred Techniques. Developed by the Described and Captioned Media Program (DCMP)**. 2014. Available: <<http://www.captioningkey.org/>>. Accessed: 29/06/2015. Cited 5 times on pages 84, 86, 142, 144, and 145.

DEMERS, J. **The top 7 online marketing trends that will dominate 2016**. 2015. Available: <<http://www.forbes.com/sites/jaysondemers/2015/09/29/the-top-7-online-marketing-trends-that-will-dominate-2016/#5777dded34c04>>. Accessed: 02/02/2016. Cited on page 23.

DEPARTAMENTO DE GOVERNO ELETRÔNICO. **eMAG - Modelo de Acessibilidade em Governo Eletrônico**. 2014. Available: <<http://emag.governoeletronico.gov.br/>>. Accessed: 18/05/2015. Cited 5 times on pages 30, 34, 35, 36, and 37.

DIAS, A. L. **Um processo para sistemas web com foco em acessibilidade e usabilidade**. 287 p. Phd Thesis (Doctoral thesis) — Instituto de Ciências Matemáticas e de Computação (ICMC), São Carlos, Universidade de São Paulo, 2014. Available: <<http://www.teses.usp.br/teses/disponiveis/55/55134/tde-18032015-160137/>>. Accessed: 02/02/2016. Cited on page 51.

DIGITALGOV. **Making Multimedia Section 508 Compliant and Accessible**. 2013. Available: <<http://www.digitalgov.gov/2013/06/26/making-multimedia-section-508-compliant-and-accessible/>>. Accessed: 30/10/2015. Cited 5 times on pages 35, 36, 90, 91, and 175.

_____. **508 Accessible Videos – How to Caption Videos**. 2014. Available: <<https://www.digitalgov.gov/2014/06/30/508-accessible-videos-how-to-caption-videos/>>. Accessed: 30/10/2015. Cited on page 33.

_____. **508 Accessible Videos – How to Make Audio Descriptions**. 2014. Available: <<https://www.digitalgov.gov/2014/06/30/508-accessible-videos-how-to-make-audio-descriptions/>>. Accessed: 30/10/2015. Cited 2 times on pages 34 and 87.

DO-IT. **How can educational entities plan an accessible video production?** 2013. Available: <<http://www.washington.edu/doi/Conf/articles?1211>>. Accessed: 29/06/2015. Cited 5 times on pages 35, 36, 77, 78, and 81.

ENCELLE, B.; OLLAGNIER-BELDAME, M.; POUCHOT, S.; PRIÉ, Y. Annotation-based video enrichment for blind people: A pilot study on the use of earcons and speech synthesis. In: **The Proceedings of the 13th International ACM SIGACCESS Conference on Computers and Accessibility**. New York, NY, USA: ACM, 2011. (ASSETS '11), p. 123–130. ISBN 978-1-4503-0920-2. Available: <<http://doi.acm.org/10.1145/2049536.2049560>>. Cited 2 times on pages 25 and 43.

FACEBOOK. **What the Shift to Video Means for Creators**. 2015. Available: <<http://media.fb.com/2015/01/07/what-the-shift-to-video-means-for-creators/>>. Accessed: 18/11/2015. Cited on page 23.

FAJARDO, M. C. d. S.; ORTENZI, D. I. B. G. **Letramento digital e midiático: desafios para o ensino da língua inglesa no século XXI**. 2007. Available: <https://ensinodelinguascomtic.files.wordpress.com/2010/03/webquest-para-letramento_digital_e_midiatico.pdf>. Accessed: 18/05/2015. Cited on page 79.

GARCÍA-CRESPO, Á.; LÓPEZ-CUADRADO, J. L.; GONZÁLEZ-CARRASCO, I. Applications and usability of interactive tv: 4th iberoamerican conference, jauti 2015, and 6th congress on interactive digital tv, ctvdi 2015, palma de mallorca, spain, october 15-16, 2015. revised selected papers. In: _____. Cham: Springer International Publishing, 2016. chap. Accessibility on VoD Platforms via Mobile Devices, p. 149–160. ISBN 978-3-319-38907-3. Available: <http://dx.doi.org/10.1007/978-3-319-38907-3_12>. Cited on page 44.

GONZÁLEZ, M.; MORENO, L.; MARTÍNEZ, P.; IGLESIAS, A. Web accessibility requirements for media players. In: CAMPOS, P.; GRAHAM, N.; JORGE, J.; NUNES, N.; PALANQUE, P.; WINCKLER, M. (Ed.). **Human-Computer Interaction – INTERACT 2011**. Springer Berlin Heidelberg, 2011, (Lecture Notes in Computer Science, v. 6949). p. 669–674. ISBN 978-3-642-23767-6. Available: <http://dx.doi.org/10.1007/978-3-642-23768-3_109>. Cited 2 times on pages 25 and 44.

GUIMARÃES, R. L. **Socially-Aware Multimedia Authoring**. Amsterdam, The Netherlands: Amsterdam: Vrije Universiteit, 2014. ISBN 978-94-6259-028-1. Cited 5 times on pages 43, 44, 45, 46, and 50.

GUIMARÃES, R. L.; CESAR, P.; BULTERMAN, D. C. Creating and sharing personalized time-based annotations of videos on the web. In: **Proceedings of the 10th ACM Symposium on Document Engineering**. New York, NY, USA: ACM, 2010. (DocEng '10), p. 27–36. ISBN 978-1-4503-0231-9. Available: <<http://doi.acm.org/10.1145/1860559.1860567>>. Cited on page 43.

HARDMAN, L.; ROSSUM, G.; BULTERMAN, D. C. **Structured Multimedia Authoring**. Amsterdam, The Netherlands, The Netherlands, 1993. Cited on page 43.

HILERA-GONZÁLEZ, J.; CAMPO-MONTALVO, E. **Guía para crear contenidos digitales accesibles: Documentos, presentaciones, vídeos, audios y páginas web**. España: Universidad de Alcalá, 2015. ISBN 978-84-16133-52-9. Cited 3 times on pages 25, 43, and 46.

HONG, R.; WANG, M.; YUAN, X.-T.; XU, M.; JIANG, J.; YAN, S.; CHUA, T.-S. Video accessibility enhancement for hearing-impaired users. **ACM Trans. Multimedia Comput. Commun. Appl.**, ACM, New York, NY, USA, v. 7S, n. 1, p. 24:1–24:19, Nov. 2011. ISSN 1551-6857. Available: <<http://doi.acm.org/10.1145/2037676.2037681>>. Cited 2 times on pages 25 and 43.

HOSSAIN, F.; ALI, M. Relation between individual and society. **Open Journal of Social Sciences**, Scientific Research, v. 2, n. 8, p. 130–137, 2014. Available: <http://file.scirp.org/Html/8-1760197_49227.htm>. Cited on page 23.

ISO9241-171. **ISO 9241-171:2008, Ergonomics of human-system interaction - Part 171: Guidance on software accessibility**. 2008. <http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=39080>. Accessed: 2014-10-08. Cited 2 times on pages 30 and 32.

ISO/IEC71. **ISO/IEC GUIDE 71:2014, Guide for addressing accessibility in standards**. 2014. <http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=57385>. Accessed: 2015-10-08. Cited 2 times on pages 29 and 30.

KIRK, D.; SELLEN, A.; HARPER, R.; WOOD, K. Understanding videowork. In: **Proceedings of the SIGCHI Conference on Human Factors in Computing Systems**. New York, NY, USA: ACM, 2007. (CHI '07), p. 61–70. ISBN 978-1-59593-593-9. Available: <<http://doi.acm.org/10.1145/1240624.1240634>>. Cited 8 times on pages 37, 38, 39, 41, 44, 45, 50, and 80.

LAB, M. P. I.; STATE, M. **Guidelines for the Development of American Sign Language Versions of Academic Test Content for K:12 Students**. 2015. Available: <<https://www.measuredprogress.org/wp-content/uploads/2015/08/GAAP-ASL-Guidelines.pdf>>. Accessed: 29/10/2015. Cited on page 94.

LAW13.146. **LEI Nº 13.146, DE 6 DE JULHO DE 2015. Institui a Lei Brasileira de Inclusão da Pessoa com Deficiência (Estatuto da Pessoa com Deficiência)**. 2015. Available: <http://www.planalto.gov.br/ccivil_03/_Ato2015-2018/2015/Lei/L13146.htm>. Accessed: 15/10/2015. Cited on page 30.

LAWTON, H. S. **Transcripts on the Web: Getting people to your podcasts and videos**. 2009. Available: <http://www.uiaccess.com/transcripts/transcripts_on_the_web.html>. Accessed: 11/11/2015. Cited 6 times on pages 35, 36, 89, 90, 91, and 175.

LI, Y.; WALD, M.; KHOJA, S.; WILLS, G.; MILLARD, D.; KAJABA, J.; SINGH, P.; GILBERT, L. Synote: Enhancing multimedia e-learning with synchronised annotation. In: **Proceedings of the First ACM International Workshop on Multimedia Technologies for Distance Learning**. New York, NY, USA: ACM, 2009. (MTDL '09), p. 9–18. ISBN 978-1-60558-757-8. Available: <<http://doi.acm.org/10.1145/1631111.1631114>>. Cited 3 times on pages 25, 44, and 47.

LOPEZ-COLINO, F.; COLAS, J. Spanish sign language synthesis system. **Journal of Visual Languages and Computing**, v. 23, n. 3, p. 121 – 136, 2012. ISSN 1045-926X. Available: <<http://www.sciencedirect.com/science/article/pii/S1045926X12000122>>. Cited 2 times on pages 25 and 43.

MEDIA; EDUCATION, U. Information Services Department Faculty of. **Introduction to Video Production**. 2012. Available: <<https://www.edu.uwo.ca/mis/handouts/video-production.pdf>>. Accessed: 07/12/2015. Cited 3 times on pages 37, 38, and 40.

MEIXNER, B.; MATUSIK, K.; GRILL, C.; KOSCH, H. Towards an easy to use authoring tool for interactive non-linear video. **Multimedia Tools Appl.**, Kluwer Academic Publishers, Hingham, MA, USA, v. 70, n. 2, p. 1251–1276, May 2014. ISSN 1380-7501. Available: <<http://dx.doi.org/10.1007/s11042-012-1218-6>>. Cited on page 43.

MINNEMAN, S.; HARRISON, S.; JANSSEN, B.; KURTENBACH, G.; MORAN, T.; SMITH, I.; MELLE, B. van. A confederation of tools for capturing and accessing collaborative activity. In: **Proceedings of the Third ACM International Conference on Multimedia**. New York, NY, USA: ACM, 1995. (MULTIMEDIA '95), p. 523–534. ISBN 0-89791-751-0. Available: <<http://doi.acm.org/10.1145/217279.215316>>. Cited on page 65.

ORGANIZATION, W. H. **Disability and health**. 2015. Available: <<http://www.who.int/mediacentre/factsheets/fs352/en/>>. Accessed: 12/02/2016. Cited on page 24.

PAE. **Guia de Accesibilidad en Contenidos Multimedia developed by Subdirección General de Información, Documentación y Publicaciones, Accesibilidad del Portal de Administración Electrónica (PAe), Dirección de Tecnologías de la Información y las Comunicaciones**. 2014. Available: <<http://administracionelectronica.gob.es/PAe/accesibilidad/documentacion>>. Accessed: 02/02/2016. Cited 3 times on pages 25, 43, and 46.

PFEIFFER, S. **The Definitive Guide to HTML5 Video**. 1st. ed. Berkely, CA, USA: Apress, 2010. ISBN 1430230908, 9781430230908. Cited on page 47.

PRODEAF. **Web Libras Avatar**. 2015. Available: <<http://weblibras.com.br/>>. Accessed: 11/10/2015. Cited on page 93.

REA, P.; IRVING, D. **Producing and Directing the Short Film and Video**. Taylor & Francis, 2015. ISBN 9781317908753. Available: <<https://books.google.com.br/books?id=7dKTBwAAQBAJ>>. Cited 2 times on pages 24 and 25.

RODRIGUES, K. R. H.; MELO, E. L.; NAKAGAWA, P. I.; TEIXEIRA, C. A. C. Interacao com conteudo complementar para apoio ao entendimento de programas televisivos. In: **Proceedings of the IX Symposium on Human Factors in Computing Systems**. Porto Alegre, Brazil, Brazil: Brazilian Computer Society, 2010. (IHC '10), p. 91–100. Available: <<http://dl.acm.org/citation.cfm?id=1999593.1999604>>. Cited 2 times on pages 25 and 43.

ROSAS-VILLENA, J. M.; RAMOS, B.; GOULARTE, R.; FORTES, R. P. Video accessibility on the most accessed websites - a case study regarding visual disabilities. In: ANTONA, M.; STEPHANIDIS, C. (Ed.). **Universal Access in Human-Computer Interaction. Access to Today's Technologies**. Springer International Publishing, 2015, (Lecture Notes in Computer Science, v. 9175). p. 231–241. ISBN 978-3-319-20677-6. Available: <http://dx.doi.org/10.1007/978-3-319-20678-3_23>. Cited 3 times on pages 24, 74, and 86.

ROWE, L. A. Looking forward 10 years to multimedia successes. **ACM Trans. Multimedia Comput. Commun. Appl.**, ACM, New York, NY, USA, v. 9, n. 1s, p. 37:1–37:7, Oct. 2013. ISSN 1551-6857. Available: <<http://doi.acm.org/10.1145/2490825>>. Cited on page 25.

SADALLAH, M.; AUBERT, O.; PRIÉ, Y. Chm: An annotation- and component-based hypervideo model for the web. **Multimedia Tools Appl.**, Kluwer Academic Publishers, Hingham, MA, USA, v. 70, n. 2, p. 869–903, May 2014. ISSN 1380-7501. Available: <<http://dx.doi.org/10.1007/s11042-012-1177-y>>. Cited on page 43.

SECTION508. **Section 508 Standards for Electronic and Information Technology**. 2000. Available: <<http://www.access-board.gov/guidelines-and-standards/communications-and-it/about-the-section-508-standards/section-508-standards>>. Accessed: 18/05/2015. Cited 3 times on pages 30, 34, and 35.

SHARDA, N. Multimedia authoring. In: FURHT, B. (Ed.). **Encyclopedia of Multimedia**. Springer US, 2008. p. 497–502. ISBN 978-0-387-74724-8. Available: <http://dx.doi.org/10.1007/978-0-387-78414-4_127>. Cited on page 43.

SOCIAL, C. **Censo demográfico 2010: características gerais da população, religião e pessoas com deficiência**. 2012. Available: <http://biblioteca.ibge.gov.br/visualizacao/periodicos/94/cd_2010_religiao_deficiencia.pdf>. Accessed: 22/06/2015. Cited 2 times on pages 24 and 37.

SUBCHEFIA PARA ASSUNTOS JURÍDICOS, CASA CIVIL, PRESIDÊNCIA DA REPÚBLICA. **LEI Nº 8.069, DE 13 DE JULHO DE 1990. Dispõe sobre o Estatuto da Criança e do Adolescente e dá outras providências**. 1990. Available: <http://www.planalto.gov.br/ccivil_03/LEIS/L8069.htm>. Accessed: 18/05/2015. Cited on page 78.

_____. **LEI No 10.741, DE 1º DE OUTUBRO DE 2003. Dispõe sobre o Estatuto do Idoso e dá outras providências**. 2003. Available: <http://www.planalto.gov.br/ccivil_03/leis/2003/L10.741.htm>. Accessed: 18/05/2015. Cited on page 78.

UNIFENAS. **Sinais de inclusão, Libras - Língua Brasileira de Sinais**. 2010. Available: <<http://pt.slideshare.net/mariavaleriab/cartilha-libras-51637265>>. Accessed: 29/10/2015. Cited on page 94.

VIGO, M.; BRAJNIK, G.; CONNOR, J. O. **Research Report on Web Accessibility Metrics**. 2012. Available: <<https://www.w3.org/TR/accessibility-metrics-report/>>. Accessed: 20/02/2016. Cited on page 95.

VILLAMIZAR, J. F. S.; ENCELLE, B.; PRIÉ, Y.; CHAMPIN, P.-A. An adaptive videos enrichment system based on decision trees for people with sensory disabilities. In: **Proceedings of the International Cross-Disciplinary Conference on Web Accessibility**. New York, NY, USA: ACM, 2011. (W4A '11), p. 7:1–7:4. ISBN 978-1-4503-0476-4. Available: <<http://doi.acm.org/10.1145/1969289.1969299>>. Cited 2 times on pages 25 and 43.

VILLENA. **Accessibilidade em vídeos**. 2014. Available: <<http://accessibilidadevideos.blogspot.com.br/>>. Accessed: 2014-12-03. Cited 3 times on pages 65, 66, and 70.

W3C. **Audio Description or Media Alternative (Prerecorded): Understanding SC 1.2.3**. 2008. Available: <<http://www.w3.org/TR/UNDERSTANDING-WCAG20/media-equiv-audio-desc.html>>. Accessed: 11/10/2015. Cited 3 times on pages 33, 35, and 90.

_____. **Captions (Prerecorded): Understanding SC 1.2.2**. 2008. Available: <<https://www.w3.org/TR/UNDERSTANDING-WCAG20/media-equiv-captions.html>>. Accessed: 18/05/2015. Cited on page 33.

_____. **Sign Language (Prerecorded): Understanding SC 1.2.6**. 2008. Available: <<http://www.w3.org/TR/UNDERSTANDING-WCAG20/media-equiv-sign.html>>. Accessed: 11/10/2015. Cited 2 times on pages 33 and 37.

_____. **Web Accessibility for Older Users: A Literature Review**. 2008. Available: <<http://www.w3.org/TR/wai-age-literature/>>. Accessed: 26/10/2015. Cited on page 78.

_____. **Web Content Accessibility Guidelines 2.0**. 2008. Available: <<http://www.w3.org/TR/2008/REC-WCAG20-20081211/>>. Accessed: 18/05/2015. Cited 6 times on pages 29, 30, 32, 33, 34, and 82.

_____. **What is Web Accessibility**. 2012. Available: <<http://www.w3.org/WAI/intro/accessibility.php>>. Accessed: 18/05/2015. Cited on page 29.

_____. **Alternative content definition by User Agent Accessibility Guidelines (UAAG) 2.0**. 2015. Available: <<http://w3c.github.io/UAAG/UAAG20/>>. Accessed: 28/12/2015. Cited 2 times on pages 32 and 33.

_____. **Authoring Tool Accessibility Guidelines (ATAG) 2.0**. 2015. Available: <<https://www.w3.org/TR/2015/REC-ATAG20-20150924/>>. Accessed: 18/10/2015. Cited on page 30.

_____. **Media Accessibility User Requirements - Transcript**. 2015. Available: <<https://www.w3.org/TR/media-accessibility-reqs/>>. Accessed: 26/01/2016. Cited on page 35.

_____. **User Agent Accessibility Guidelines (UAAG) 2.0**. 2015. Available: <<http://www.w3.org/TR/2015/WD-UAAG20-20150915/>>. Accessed: 18/11/2015. Cited 3 times on pages 29, 30, and 32.

W3SCHOOLS. **HTML5 Video**. 2015. Available: <http://www.w3schools.com/html/html5_video.asp>. Accessed: 29/06/2015. Cited on page 82.

WILLIAMS, A. User-centered design, activity-centered design, and goal-directed design: A review of three methods for designing web applications. In: **Proceedings of the 27th ACM International Conference on Design of Communication**. New York, NY, USA: ACM, 2009. (SIGDOC '09), p. 1–8. ISBN 978-1-60558-559-8. Available: <<http://doi.acm.org/10.1145/1621995.1621997>>. Cited on page 53.

YIN, R. **Case Study Research: Design and Methods**. SAGE Publications, 2009. (Applied Social Research Methods). ISBN 9781412960991. Available: <<https://books.google.com.br/books?id=FzawIAdilHkC>>. Cited on page 107.

YOUTUBE. **Estatísticas**. 2015. Available: <<https://www.youtube.com/yt/press/pt-BR/statistics.html>>. Accessed: 18/05/2015. Cited on page 23.

GLOSSARY

Accessibility: refers to the extent to which products, systems, services, environments and facilities can be used by people from a population with the widest range of characteristics and capabilities to achieve a specified goal in a specified context of use.

Alternative content: is the content as text (caption and transcript), audio (audio description) and video (sign language); needed to understand the media content, synchronized with the video.

Audience: is a group of people who watch the videos.

Audio description: is an additional audio track to describe important visual details that cannot be understood from the main soundtrack alone. AD provides information about actions, characters, scene changes, on-screen text, and other visual content. The narration is added during existing pauses in dialogue.

Closed Caption: only appear when is turn then on, it can be turn them off. Include not only dialogue, but also non-speech information such as speaker identification, music, laughter and meaningful sound effects.

Footage: is the unedited material, recorded by a some device, which usually must be edited to create a final product.

Libras: Braziliam Sign Language.

Transcript: is a media alternative for synchronized media. Its purpose is to provide a description of the video of all visual information, including visual context, actions and expressions of actors, and any other visual material.

ACCESSIBILITY REQUIREMENTS FOR ALTERNATIVE CONTENT

The accessibility requirements were developed from the guidelines and recommendations of the alternative content explained before at Alternative content authoring. The requirements related to functionalities of video player were excluded. After excluding, it was selected 109 requirements for captions, 75 requirements for audio description and 28 requirements for transcript. Following are described the accessibility requirements for captions (see Table 18), audio description (see Table 19) and transcript (see Table 20).

Each table has the following structure: a number of the question, accessibility guideline description, level, applicability, conformance and comments. The level is filled with M (Mandatory) or R (Recommended). The applicability is filled with Y (Yes) and N (No).

The specialist should verify if the accessibility requirements is applicable to the alternative content that is being evaluated for a specific video. The conformance is filled with Y (Yes) and N (No). If the specialist determined that an accessibility requirement is applied to the video, it is necessary verify if the accessibility requirement was addressed or not for a specific video.

Table 18 – Requirements for caption

Nro	Guideline	Level	Applicability		Conformance	
		M, R	Y,N	Comments	Y,N	Comments
Q1	Do all text (case, font, line division, caption duration, and caption placement) have a correct appearance and presentation?					
1.1	Mixed case characters are preferred for readability. However, capital letters are used for screaming or shouting.	R				
1.2	When a sentence is broken into two or more lines of captions, it should be broken at a logical point where speech normally pauses.	M				
1.3	Do not break a modifier from the word it modifies.	M				
1.4	Do not break a prepositional phrase.	M				
1.5	Do not break a person’s name nor a title from the name with which it is associated.	M				
1.6	Do not break a line after a conjunction.	M				
1.7	Do not break an auxiliary verb from the word it modifies.	M				
1.8	Never end a sentence and begin a new sentence on the same line unless they are short, related sentences containing one or two words.	M				
1.9	Captions should have a minimum length of 40 frames (1 second and 10 frames).	M				
1.10	Captions should have a maximum length of 6 seconds. (Background music notation is an exception to this guideline).	M				
1.11	It is preferred that there are no more than two lines per caption.	R				
1.12	Three- or four-line captions are occasionally acceptable if a one-or two-line caption would interfere with preexisting graphics or be confused regarding the speaker identification.	R				
Continued on next page						

Table 18 – continued from previous page

Nro	Guideline	Level	Applicability		Conformance	
		M, R	Y,N	Comments	Y,N	Comments
Q2	Do captions incorporate the proper use of spelling, grammar, punctuation, capitalization, and other factors deemed necessary for high-quality captioned media?					
2.1	Do not emphasize a word using all capital letters except to indicate screaming or shouting.	M				
2.2	Be consistent in the spelling of words throughout the media. This includes vocabulary that can be spelled either as one or two words or in hyphenated form.	M				
2.3	Capitalize proper nouns for speaker identification. All other speaker identification should be lowercased unless this identification is being used as a proper noun.	M				
2.4	When captioning a list separated by commas, use a serial, or Oxford, comma.	M				
2.5	Lowercase sound effects, including both description and onomatopoeia, except when a proper noun is part of the description.	M				
2.6	When a speaker hesitates or stutters, caption what is said. Separate syllable or letter with hyphens.	M				
2.7	When captioning spelling (including fingerspelling), separate capital letters with hyphens.	M				
2.8	Use an ellipsis when there is a significant pause within a caption.	M				
2.9	Do not use an ellipsis to indicate that the sentence continues into the next caption.	M				
2.10	Use an ellipsis to lead into or out of audio related to an onscreen graphic.	M				
Continued on next page						

Table 18 – continued from previous page

Nro	Guideline	Level	Applicability		Conformance	
		M, R	Y,N	Comments	Y,N	Comments
2.11	Use quotation marks for onscreen readings from a poem, book, play, journal, or letter. However, use quotation marks and italics for offscreen readings or voice-overs.	M				
2.12	Beginning quotation marks should be used for each caption of quoted material except for the last caption. The last caption should have only the ending quotation mark.	M				
2.13	Spaces should not be inserted before ending punctuation, after opening and before closing parentheses and brackets, before and after double hyphens and dashes, or before/between/after the periods of an ellipsis.	M				
2.14	A space should be inserted after the beginning music icon (§) and before the ending music icon(s).	M				
2.15	Use italics for a voice-over reading of a poem, book, play, journal, letter, etc. (This is also quoted material, so quotation marks are also needed.)	R				
2.16	Use italics when a person is dreaming, thinking, or reminiscing.	R				
2.17	Use italics when there is background audio that is essential to the plot, such as a PA system or TV.	R				
2.18	Use italics for the first time a new word is being defined, but do not italicize the word thereafter.	R				
2.19	Use italics for offscreen dialogue, narrator, sound effects, or music (this includes background music). If there is only one person speaking throughout the program (including the narrator), whether onscreen or offscreen, use Roman type with no italics.	R				
Continued on next page						

Table 18 – continued from previous page

Nro	Guideline	Level	Applicability		Conformance	
		M, R	Y,N	Comments	Y,N	Comments
2.20	Use italics for the offscreen narrator when there are multiple speakers onscreen or offscreen.	R				
2.21	Use italics for speaker identification when the captioned dialogue is in italics.	R				
2.22	Use italics for foreign words and phrases, unless they are in an English dictionary.	R				
2.23	Use italics when a particular word is heavily emphasized in speech.	R				
2.24	When an entire caption is already in italicized format, use Roman type to set off a word that would normally be italicized.	R				
2.25	Do not italicize when a person who is offscreen is translating for a speaker who is onscreen.	R				
Q3	Is the presentation rate adequate?					
3.1	All lower-level educational media should optimally be captioned at a presentation rate not to exceed 130 words per minute (wpm), middle-level not to exceed 140 wpm, and upper-level not to exceed 160 wpm.	M				
3.2	Presentation rate is ignored only when any person is quoted, a well-known or famous person is speaking on-screen or off-screen, poems and other published works are quoted, and/or song lyrics are sung. These must be captioned verbatim.	M				
3.3	Editing is performed only when a caption exceeds a specified presentation rate limit. One way of achieving the desired presentation rate is by borrowing 15 frames before or after the audio occurs. This "borrowing" technique is hardly noticeable to the viewer.	R				
Continued on next page						

Table 18 – continued from previous page

Nro	Guideline	Level	Applicability		Conformance	
		M, R	Y,N	Comments	Y,N	Comments
3.4	Editing can be relatively simple by eliminating redundant and/or nonessential information.	R				
3.5	Editing becomes more complex when it is necessary to move beyond the simple elimination of non-essential information and may involve shortening, deleting, or reordering words, phrases, or sentences. Such editing must keep the original meaning, content, and essential vocabulary.	R				
Q4	Are sound effects (music, narration, or dialogue) correctly described?					
4.1	A description of sound effects should include the source of the sound, it can be made in two ways: (1) in brackets (DCMP, 2014), or (2) use parentheses and uppercase the first letter (CAB, 2012).	M				
4.2	Described sound effects may be combined with onomatopoeia. The described sound effect should be on the first line of the caption, separate from the onomatopoeia. Both described sound effects and onomatopoeias must be lowercased.	R				
4.3	Offscreen sound effects should be italicized, if italics are available.	R				
4.4	Place the description of the sound effect as close as possible to the sound source.	R				
4.5	For offscreen sound effects, it is not necessary to repeat the source of the sound if it is making the same sound a few captions later.	R				
4.6	Use punctuation to indicate speed or pace of sound.	M				
4.7	A sound represented by a repeated word is not hyphenated. A sound represented by two different words is hyphenated.	M				
Continued on next page						

Table 18 – continued from previous page

Nro	Guideline	Level	Applicability		Conformance	
		M, R	Y,N	Comments	Y,N	Comments
4.8	When describing a sustained sound, use the present participle form of the verb. When describing an abrupt sound, use the third person verb form.	M				
4.9	Caption background sound effects only when they are essential to the plot.	M				
4.10	When it is possible, use specific rather than vague, general terms to describe sounds.	M				
4.11	Do never use the past tense when describing sounds. Captions should be synchronized with the sound and are therefore in the present tense.	M				
4.12	A description should be used for instrumental/background music when it is essential to the understanding of the program.	M				
4.13	Offscreen background music description should be italicized.	R				
4.14	If it's possible, the music description should include the performer/composer and the title.	R				
4.15	Use descriptions that indicate the music mood. Be as objective as possible. Avoid subjective words, such as "delightful," "beautiful," or "melodic."	M				
4.16	Beware of misplaced modifiers in music descriptions.	R				
4.17	Nonessential background music should be captioned by placing a music icon (§) on the upper right corner of the screen and it should never be captioned at the expense of dialogue.	R				
4.18	Do not caption background music which is under 5 seconds.	M				
4.19	If music contains lyrics, caption the lyrics verbatim. The lyrics should be introduced with the name of the artist and the title in brackets, if the presentation rate permits.	R				
Continued on next page						

Table 18 – continued from previous page

Nro	Guideline	Level	Applicability		Conformance	
		M, R	Y,N	Comments	Y,N	Comments
4.20	Caption lyrics with music icons (§). Use one music icon at the beginning and one at the end of each caption within a song, but use two music icons at the end of the last line of a song. A space should be inserted after the beginning music icon (§) and before the ending music icon(s).	M				
Q5	Is speaker's identification correctly specified?					
5.1	When possible, use caption placement to identify an onscreen speaker by placing the caption under the speaker.	R				
5.2	Do not identify the speaker by name until the speaker is introduced in the audio or by an onscreen text/graphic.	M				
5.3	When a speaker cannot be identified by placement and his/her name is known, there are two possibilities: (1) the speaker's name should be in parentheses (DCMP, 2014) or (2) capitalize proper nouns and use colon at the end (CAB, 2012).	M				
5.4	The speaker's name needs to be on a line of its own, separate from the captions.	M				
5.5	When a speaker cannot be identified by placement and his/her name is unknown, identify the speaker using the same information available to the hearing viewer.	M				
5.6	If a speaker is offscreen and his/her offscreen position is known, place the captions to the far right or left, as close as possible onscreen to the offscreen speaker's assumed position.	R				
Continued on next page						

Table 18 – continued from previous page

Nro	Guideline	Level	Applicability		Conformance	
		M, R	Y,N	Comments	Y,N	Comments
5.7	If there are multiple speakers and only one narrator, identify as (female narrator) or (male narrator) at the beginning of the media. It is not necessary to identify gender for each caption thereafter.	M				
5.8	When an actor is portraying another person or character, identify the actor as the person being portrayed.	M				
Q6	Are captions correctly synchronized?					
6.1	Captions should closely match the original audio. Keeping the textual unity with picture and sound ensures clarity.	M				
6.2	Do not simultaneously caption different speakers if they are not speaking at the same time.	M				
Q7	Are the intonation, play on words and no audio considered?					
7.1	If the speaker is not visible onscreen, or visual clues that denote the emotional state are not shown, indicate the speaker’s emotion.	M				
7.2	When a person is whispering, caption as: [wispering] (DCMP, 2014).	M				
7.3	When feasible, describe puns.	M				
7.4	When people are seen talking, but there is no audio, caption as [no audio] or [silence] (DCMP, 2014).	M				
Q8	Are foreign language, dialect, slang and phonetics considered?					
8.1	If possible, caption the actual foreign words. If it is not possible to caption the words, use a description (e.g., [speaking French]). Do never translate into English.	R				
Continued on next page						

Table 18 – continued from previous page

Nro	Guideline	Level	Applicability		Conformance	
		M, R	Y,N	Comments	Y,N	Comments
8.2	If possible, use accent marks, diacritical marks, and other indicators.	R				
8.3	Indicate regional accent at the beginning of the first caption.	M				
8.4	Keep the flavor of dialect.	M				
8.5	Caption profanity and slang if in the audio.	M				
8.6	When a word is spoken phonetically, caption it the way it is commonly written.	M				
Q9	Are numbers (dates, time, periods of time, fractions, percentages, amounts, and measurements) correctly described?					
9.1	Spell out all numbers from one to ten, but use numerals for all numbers over ten.	M				
9.2	Spell out any number that begins a sentence as well as any related numbers.	M				
9.3	Spell out casual, no emphatic numbers.	M				
9.4	Four-digits numbers can either have a comma or not. Be consistent throughout the media production. For numerals having over five digits, a comma is necessary.	M				
9.5	Use numerals when referring to technical and athletic terms.	M				
9.6	Use numerals in a listing of numbers if one or more is above ten and these occur in one caption or one sentence.	M				
9.7	When indicating sequence, capitalize the noun and use numerals. Exceptions are the indication of line, note, page, paragraph, size, step, or verse.	M				
9.8	Use the numeral plus the lowercase "th," "st," or "nd" when a day of the month is mentioned by itself (no month is referred to).	M				
Continued on next page						

Table 18 – continued from previous page

Nro	Guideline	Level	Applicability		Conformance	
		M, R	Y,N	Comments	Y,N	Comments
9.9	When the day precedes the month, use the numeral plus the lowercase "th," "st," or "nd" if the ending is spoken.	M				
9.10	Use the numeral alone when the day follows the month.	M				
9.11	When the month, day, and year are spoken, use the numeral alone for the day, even if an ending ("th," "st," or "nd") is spoken.	M				
9.12	Indicate time of day with numerals only.	M				
9.13	Always use numerals when the abbreviation "a.m." or "p.m." is present. Double zeros are not necessary to indicate minutes of the hour when a whole number is used with a.m. or p.m.	M				
9.14	A decade should be captioned as "the 1980s" (not "the 1980's") and "the '50s" (not "the 50's").	M				
9.15	If a decade or century is in noun form, do not use hyphens.	M				
9.16	If a period of time is used as an adjective, use a hyphen.	M				
9.17	Either spell out or use numerals for fractions, keeping this rule consistent throughout the media. If using numerals, insert a space between a whole number and its fraction.	M				
9.18	Do not mix numerals and spelled-out words within the same sentence.	M				
9.19	If a fraction is used with "million," "billion," "trillion," etc., spell out the fraction.	M				
9.20	Fractions expressed in figures should not be followed by endings, such as "sts," "rds," "nds," or "ths."	M				
Continued on next page						

Table 18 – continued from previous page

Nro	Guideline	Level	Applicability		Conformance	
		M, R	Y,N	Comments	Y,N	Comments
9.21	Use numerals and the percent sign to indicate all percentages except at the beginning of a sentence.	M				
9.22	Use the numeral plus "cents" or its symbol for amounts under one dollar. Use the equivalent for other currency.	M				
9.23	Use the dollar sign plus the numeral for dollar amounts under one million. For whole-dollar amounts of one million and greater, spell out "million," "billion," etc. Use the equivalent for other currency.	M				
9.24	Use the word "dollar" only once for a range up to ten. Use the equivalent for other currency.	M				
9.25	Use the dollar sign and numerals when captioning a range of currency over ten dollars. Use the equivalent for other currency.	M				
9.26	Spell out units of measurement, such as "inches," "feet," "yards," "miles," "ounces," "pounds," and "tablespoons." However, if spoken in shortened form, symbols should be used.	M				
9.27	For whole numbers, use numerals. For example, caption "2 cups of sugar" instead of "two cups of sugar."	M				

Table 19 – Requirements for audio description

Nro	Guideline	Level	Applicability		Conformance	
		M, R	Y,N	Comments	Y,N	Comments
Q1	Does description describe what you see?					
1.1	What you see is what you describe. One sees physical appearances and actions; one does not see motivations or intentions. Do never describe what you think you see.	M				
1.2	Preview the material bearing in mind that it should include the visual information that is inaccessible to people who are blind or have low vision. It includes key plot elements, people, places, actions, objects, unknown sound sources, etc., not mentioned in the dialogue or made obvious by what one hears. Concentrate on what is the most significant and least obvious from the dialogue or other audio information. Describing everything is impossible—so, describe what is essential on the available time.	M				
1.3	Describe essentials first and then, as time permits, describe further elements such as the decorative details of the setting, the physical appearance and mannerism of the characters, architecture, clothing style, technology, color, light and texture. This sort of description works well during long pauses in the action or during scene changes.	M				
1.4	Description should not fill every available pause. Less is more. Audio description is not a running commentary. Listeners should be allowed to hear the emotion in actors' voices and in the tension of the silences between characters.	M				
1.5	Be sure to describe, as casually as possible, seemingly insignificant things the sighted audience will observe without knowing their later importance.	M				
Continued on next page						

Table 19 – continued from previous page

Nro	Guideline	Level	Applicability		Conformance	
		M, R	Y,N	Comments	Y,N	Comments
1.6	Most audio description for film and video is scripted and recorded on an audio track separate from the material’s soundtrack. This process allows the describer to write complete, accurate descriptions that will precisely fit the available pauses. Be sure to read the script aloud at the rate it will be read for recording to verify its timing and to check for tongue-twisters.	M				
1.7	The material’s producer should approve the descriptions before the recording session - with the understanding that there will almost certainly be minor script adjustments to refine the timing during the recording.	M				
1.8	Because pauses in the material are often brief, use concise phrases or sentences. Shorter descriptions are also easier for listeners to grasp.	M				
1.9	If a description is essential and a silence is especially short, the describer may have to step on the first syllable or two of dialogue or narration. This often occurs when the “next voice” must be identified so listeners will understand the speaker’s interest point.	M				
Q2	Does description describe objectively?					
2.1	Allow listeners to form their own opinions and draw their own conclusions. Do not editorialize, interpret, explain, analyze or help listeners in any other way.	M				
2.2	If the conclusion is that a character is angry, describe what led to that conclusion, the gestures/facial expressions of the character. Character’s moods, motives or reasoning are not visible and, thus, not subject to description.	M				
Continued on next page						

Table 19 – continued from previous page

Nro	Guideline	Level	Applicability		Conformance	
		M, R	Y,N	Comments	Y,N	Comments
2.3	Use only those adjectives and adverbs that do not offer value judgments and that are not themselves subject to interpretation.	M				
2.4	It is more interesting to name the items in the clutter if time allows to do so: “The attic is cluttered.” Do not take a series of specific, separate actions/events/images and describe them as one.	R				
2.5	When describing sizes, round-off to the next logical increment to give listeners numbers that are easier to hear and comprehend.	M				
2.6	Use the first person when the director has created a “first person” point of view as a means of including the audience.	M				
Q3	Is the description allowing the listeners to hear the dialogue?					
3.1	Listeners want to hear the performance first and the description later.	M				
3.2	A describer may talk over background music or underscoring as well as the lyrics of a repeated chorus of a song. One should not describe during significant arias, the verse of a song or its first chorus.	M				
3.3	The dialogue from radio, television or other speaking characters may be important to the story or may be considered background sound. If it is background noise, it is possible to describe over it, assuming the description is vital.	R				
3.4	Don’t talk in paragraphs or elaborate sentences. Use short phrases instead of full sentences. Try to speak at least two or three words so listeners have the opportunity to switch focus to the describer’s voice.	R				
Continued on next page						

Table 19 – continued from previous page

Nro	Guideline	Level	Applicability		Conformance	
		M, R	Y,N	Comments	Y,N	Comments
3.5	If the musical score for a film or video contributes significantly to the material's emotion, try to allow listeners to be carried along with the music without interjecting descriptions. Only interrupt for vital information that must be described during the music in order to be timely.	R				
Q4	Does description allow listeners to comprehend the material?					
4.1	Trust listeners, who, in most instances have made the choice to attend the performance, to grasp the meaning of the material and the description. Do not condescend, patronize, or talk down to listeners.	M				
4.2	If the play has a complex plot and/or a confusing set of characters, there's probably a synopsis in the playbill. Just as this information will be helpful to sighted audience members, sharing this information with listeners during pre-show notes may aid their appreciation of the performance and the description.	R				
4.3	The describer has the responsibility to find the median description, between minimal description and essential clarifications.	R				
Q5	Does description describe the material without censorship?					
5.1	Describers must say the factual information about nudity, sexual acts, violence, etc. Listeners should know everything that is evident to sighted people.	M				
Q6	Is language consistent?					
6.1	Choose language that is consistent with the content of the material.	M				
Continued on next page						

Table 19 – continued from previous page

Nro	Guideline	Level	Applicability		Conformance	
		M, R	Y,N	Comments	Y,N	Comments
6.2	Use language appropriate for the listeners. Children’s programs should use vocabulary suitable for the age group. Make every effort to pronounce words properly.	M				
6.3	Once the describer establishes a name for characters, places, objects, etc., always use that same name.	M				
6.4	Avoid metaphors, similes, etc.; however, common comparisons that should be recognized by most listeners are acceptable as a means of saving time.	M				
6.5	Not all listeners will understand slang, colloquialisms, and regional terms. Use within context of the performance.	M				
6.6	Use the correct terminology so long as the majority of listeners will understand that terminology.	M				
6.7	Use the most descriptive words and concise phrase or sentence structure.	M				
6.8	Use vivid verbs. People frequently “walk” but they also amble, stagger, shuffle, saunter, and stroll. Choose the word that best matched the action.	M				
6.9	Use pronouns carefully. If there is only one female in a scene, then “she” is fine. If there is more than one, proper names will be clearer.	M				
6.10	Address time shifts (flash backs or visions of the future) in relation to the character. Music and visual effects may further identify time changes.	M				
6.11	Use “while” and “as” to join two actions only if there is a connection between them.	M				
6.12	Describe colors both to help people with low vision to locate what’s being described and to share the emotional “meaning” of the color in the production.	M				
Continued on next page						

Table 19 – continued from previous page

Nro	Guideline	Level	Applicability		Conformance	
		M, R	Y,N	Comments	Y,N	Comments
Q7	Does description describe race, ethnicity and nationality?					
7.1	Sighted audience members do not see a character’s race, ethnicity or nationality; rather, they see skin color and facial features.	R				
7.2	When describing appearance, describers should avoid offering their judgment about a character’s cultural background.	M				
7.3	Describing skin color and facial features is an all or nothing proposition. If it’s important to note this for one character, then describe this for all characters.	R				
7.4	Rely on factual, clear, straightforward, commonly used terms that the majority of listeners will understand. Avoid disrespectful, derogatory or condescending terms as well as vague, poetic or euphemistic language. Do not use the terms “light-skinned” or “dark-skinned”. Use rules associated with the country.	M				
7.5	In a dramatic in which, perhaps because of the story and its setting, the characters’ race, ethnicity, nationality is largely apparent to sighted audience members and essential to the plot, describe the characters’ characteristics as part of the description would be helpful.	R				
Continued on next page						

Table 19 – continued from previous page

Nro	Guideline	Level	Applicability		Conformance	
		M, R	Y,N	Comments	Y,N	Comments
7.6	Plays and full-length films or videos allow time for audiences to develop an understanding of characters, including - when these classifications matter to the audiences' understanding of the material- their race, ethnicity or nationality. In other situations, such as brief interviews in video clips, visuals in museums and exhibits, etc., the listener's opportunity to learn about an individual is usually quite brief. If the subject matter deals with race/ethnicity/nationality or there are unexpected inclusions of people from multiple racial/ethnic/national backgrounds, listeners will want to know this specific information about individuals' cultural heritage rather than the observable skin color and facial features.	R				
7.7	If signed people will presume the race, ethnicity and nationality of a person and, thus, have a sense of "where the person's coming from," the describer should share this information with listeners. Whether the visuals or comments seem to follow racial expectations, sighted visitors are aware of the race of the person or persons depicted. Audio description listeners should have the same information.	M				
7.8	Check with directors, producers, etc. to determine the correct information rather than making what might be ill-founded assumptions.	R				
Q8	Are description describing from the listeners' perspective?					
8.1	Surprises should ideally come at the same time for all audience members. If characters' appearances or actions, hidden identities, costumes, sight gags, sound effects, etc. happen as a surprise to sighted audience members, don't spoil the surprise for listeners by describing (and revealing) them in advance.	M				

Continued on next page

Table 19 – continued from previous page

Nro	Guideline	Level	Applicability		Conformance	
		M, R	Y,N	Comments	Y,N	Comments
8.2	If the action that accompanies a sound effect will result in a reaction from the audience, treat this as if describing a sight gag. Time the description to allow listeners to react at the same time as sighted audience members.	M				
Q9	Are techniques applied in the description?					
9.1	Describe entrances and exits—who and where—especially when there’s nothing audible to indicate someone has joined or left the scene.	R				
9.2	Use a character’s name only when sighted audience members know the name.	M				
9.3	Establish and use one consistent name for each character as early as possible.	M				
9.4	If the characters are referred to by difficult names, be sure to use the one name that best serves the storytelling.	M				
9.5	Once the material has identified a character, match the character’s name with the actor’s voice by mentioning the character’s name just before he/she speaks.	M				
9.6	Allow the material itself to provide information to listeners.	R				
9.7	Guide people with low vision by stating the location where they should focus.	R				
9.8	Match vocal delivery to the pace, energy and volume of the material. The language and delivery to describe a fight scene would differ from that used to describe a love scene.	R				
9.9	If possible, hold description until the audience begins to quiet. If not possible, speak loudly when describing over loud laughter, music or applause.	R				
Continued on next page						

Table 19 – continued from previous page

Nro	Guideline	Level	Applicability		Conformance	
		M, R	Y,N	Comments	Y,N	Comments
9.10	When a effect will be repeated, try to describe it the first time in a way that allows a “shorthand” reference later.	R				
9.11	The narrator voice should complement the material - it should be distinct from the voices of the characters and/or the program’s narrator and mixed to sound as natural to the work as possible.	M				
9.12	If there is a great deal of dialogue or narration on the soundtrack and the film or video will be shown live rather than broadcast, the producer may agree to pre-show description during the less important beginning credits or during an added period of pre-show black screen.	R				
9.13	With a series of rapidly changing images, a technique sometimes used to establish locale or mood, briefly describe all views if possible. If not, describe the images that are most important to the plot or subject and the most visually distinguishable images to help guide people who have low vision.	M				
9.14	Preview video the first time with no picture to help identify auditory confusions-sounds that could be mistaken for something they are not, characters with similar voices, silent changes of scenes, etc.	R				
9.15	Establish the location before describing anything else.	M				
9.16	Characters in film and video may be introduced but unnamed. Use a significant physical characteristic to identify them in descriptions.	R				
Continued on next page						

Table 19 – continued from previous page

Nro	Guideline	Level	Applicability		Conformance	
		M, R	Y,N	Comments	Y,N	Comments
9.17	Without the extended time for pre-show notes, use the most revealing information about a character in place of a complete description. The age of the character may be reasonably judged by the sound of their voice. If that is not the case, it's important to mention their age. If they are dressed differently than other characters, that might be appropriate. Are they much taller or shorter than everyone else? Are they the only blonde? Do they have blue eyes, when everyone else in the family has brown?	R				
9.18	The relationships between characters may not be apparent, but it's the film maker's responsibility to reveal these relationship - not the describer's.	R				
9.19	Scene changes can be confusing particularly when the soundtrack does not indicate a change. Keep them simple and short. "In the bedroom," "at the police station," etc.	R				
9.20	As well, actions, characters, and details can be confusing if we do not know where we are. When there's a change of place, start the description with the location, even when the resulting sentence seems awkwardly structured - the information will be well structured for comprehension.	M				
9.21	Just a describer for a live performance should avoid theatrical jargon or references, which would draw listeners' attention away from their involvement in the material. A film or video describer should avoid calling attention to filmmaking errors such as mistakes in continuity or editing for the same reason.	R				
9.22	Although one should generally avoid filmmaking jargon and describing filmmaking techniques, sometimes the brevity and simplicity of something like, "The screen fades to black" is appropriate.	R				
Continued on next page						

Table 19 – continued from previous page

Nro	Guideline	Level	Applicability		Conformance	
		M, R	Y,N	Comments	Y,N	Comments
9.23	Sometimes a describer may use the word “now” to indicate a change of scene in the middle of a segment of description. Because there will be many opportunities that seem to call for the use this word, use it only when absolutely necessary.	R				
9.24	Sometimes a describer must use a silence to describe both what’s onscreen now and what’s about to appear because there’s no silence for the later information when it does appear.	R				
9.25	Occasionally there’s no silent opportunity to describe something essential to listeners’ understanding while that “something” is on the screen. The describer may need to omit a less significant description of what’s onscreen in order to interject the critical description.	R				
9.26	Describe the point of view when appropriate - “from above,” “from space,” “moving away,” “flying low over the sandy beach,” etc.	R				
9.27	We have established that we are watching film or video, so repeated references to the screen are unnecessary.	M				
9.28	Treat logos as any other image to be described and be sure to read the company name(s).	M				
9.29	Reading disclaimers and credits at the beginning and end of films, videos and television programs is an important function of audio description. As well, the describer should read text and subtitles.	M				

Table 20 – Requirements for transcript

Nro	Guideline	Level	Applicability		Conformance	
		M, R	Y,N	Comments	Y,N	Comments
Q1	Does transcript heading include video information?					
1.1	Add the type of video content, title and data. Type of video can be presentation, podcast, documentary, cooking recipes, etc.	M				
1.2	Add the abstract to made a short explanation of video content, it also could contain where the video was recorded, and who the main actors are (in case of presentations, who is presenting).	M				
1.3	If the webpage of the transcript is different from the webpage of the video, the video webpage should has a link to transcript webpage and also the transcript webpage should has a link to the video webpage.	R				
1.4	Resources as video, audio or slides should be available for download.	R				
1.5	Add a title “Transcript” before start the transcript.	R				
1.6	Use a blank line to force the start of a new description.	R				
1.7	Separate paragraphs by actors. If the video has more than one actor, the paragraphs are separated by actor. Each paragraph should contain all speech of that actor.	R				
Q2	Are speakers’ names included?					
2.1	Include speakers names.	M				
2.2	Capitalize proper nouns and use colon at the end.	M				
2.3	For an informal podcast transcript with multiple speakers, it’s probably best to use speakers’ full names the first time, and then just their first names.	M				
Continued on next page						

Table 20 – continued from previous page

Nro	Guideline	Level	Applicability		Conformance	
		M, R	Y,N	Comments	Y,N	Comments
2.4	When there are multiple speakers, use hanging indents to make it easy to skim for a particular speaker. It is possible use CSS and give each speaker a different color or style. (Make sure to include names so there is redundant color coding.)	R				
2.5	When the focus should be on the interviewee’s answers and not the interviewer, format the interviewers questions in smaller, lighter text so the interviewee’s answers stand out more clearly.	R				
Q3	Is all speech content included?					
3.1	Include all speech content.	M				
3.2	If there is speech that is not relevant, it is usually best to indicate that it has been excluded from the transcript.	R				
3.3	All additional information should be in brackets.	M				
Q4	Is relevant non-speech audio included?					
4.1	Include relevant non-speech audio.	M				
4.2	Non-speech audio is related with the use of irony, music and scenery.	M				
Q5	Is relevant visual information included?					
5.1	Visual information include visual context, actions and expressions of actors, and any other relevant visual information (photographies, unusual dress).	M				
5.2	For presentations, instead of pointing to a slide and describe, “as you can see on this slide, the traffic peaked here”, describe, “this chart of website traffic for the last year shows that it peaked in August.”	R				
Continued on next page						

Table 20 – continued from previous page

Nro	Guideline	Level	Applicability		Conformance	
		M, R	Y,N	Comments	Y,N	Comments
5.3	If the presenter does show-of-hands (e.g., “How many people follow WCAG 2.0?”), describe the results for the audio recording (e.g., “about half”).	R				
5.4	For presentations, repeat questions that are not picked up by the audio recording.	R				
Q6	Does transcript have proper punctuation and acronyms definitions?					
6.1	Check the transcripts for proper punctuation	M				
6.2	If the person speaking uses any acronyms, define them.	M				
Q7	Does transcript have an adequate edition?					
7.1	In some cases, such as legal depositions, the transcript must be verbatim, including ums, ahs, and indicating pauses.	R				
7.2	For most podcasts, presentations, and such, minor edits for readability are appropriate.	R				
7.3	It is not appropriate to change the meaning from the audio. Likewise, it is usually not appropriate to significantly correct grammar or other mistakes.	M				
Q8	Does transcript allow navigation and clarifications?					
8.1	Add headings and links where it will make the transcript more usable. For example, add links when the video is short and explains many concepts, create links for these concepts; or add titles when the video is long.	R				
8.2	Add clarifying information, as long as it is clear that it is not part of the actual audio, e.g., words added to a paragraph put in [brackets], or separate sections with headings [Introduction], [Transcript], [Resources].	R				

VIDEO4ALL METHOD, CASE STUDY RESULTS

A case study with four authors was conducted. Following are described the tables with the results of the evaluation.

Author 1 - video 1 - animation video - Table [21](#).

Author 1 - video 2 - documentary video - Table [22](#).

Author 2 - video 1 - documentary video - Table [23](#).

Author 2 - video 2 - animation video - Table [24](#).

Author 3 - video 1 - documentary video - Table [25](#).

Author 3 - video 2 - animation video - Table [26](#).

Author 4 - video 1 - animation video - Table [27](#)).

Author 4 - video 2 - documentary video - Table [28](#))

Table 21 – Author 1, video 1 (animation video)

Questions	Requirements applied	Requirements accomplished
Caption		
Q1	1.1 (R), 1.8 (M), 1.9 (M), 1.10 (M)	1.1, 1.8, 1.9, 1.10
Q2	2.1 (M), 2.2 (M), 2.3 (M), 2.6 (M), 2.9 (M), 2.13 (M)	2.1, 2.2, 2.3, 2.9, 2.13
Q3	3.1 (M)	3.1
Q4	4.1 (M), 4.2 (R), 4.7 (M), 4.10 (M), 4.11 (M)	4.1, 4.2, 4.10, 4.11
Q5	5.2 (M), 5.3 (M), 5.4 (M), 5.5 (M)	5.2, 5.3
Q6	6.1 (M), 6.2 (M)	6.1, 6.2
Q7	7.2 (M)	-
Q8	8.3 (M)	-
Q9	-	-
Audio description		
Q1	1.1 (M), 1.4 (M), 1.6 (M), 1.8 (M)	1.1, 1.4, 1.6, 1.8
Q2	2.1 (M), 2.2 (M), 2.3 (M)	2.2, 2.3
Q3	3.2 (M), 3.4 (R)	3.2, 3,4
Q4	4.1 (M), 4.3 (R)	4.3
Q5	-	-
Q6	6.1 (M), 6.6 (M), 6.8 (M), 6.9 (M)	6.1, 6.6, 6.9
Q7	7.1 (R), 7.2 (M), 7.3 (R)	7.1
Q8	-	-
Q9	9.9 (R), 9.15 (M), 9.16 (R), 9.26 (R)	9.9, 9.15, 9.16, 9.26
Transcript		
Q1	1.1 (M), 1.2 (M), 1.5 (R), 1.6 (R), 1.7 (R)	1.1, 1.2, 1.5, 1.6
Q2	2.1 (M), 2.2 (M)	2.1, 2.2
Q3	3.1 (M), 3.3 (M)	3.1, 3.3
Q4	4.1 (M), 4.2 (M)	
Q5	5.1 (M)	5.1
Q6	6.1 (M)	6.1
Q7	-	-
Q8	-	-

Table 22 – Author 1, video 2 (documentary video)

Questions	Requirements applied	Requirements accomplished
Caption		
Q1	1.1 (R), 1.8 (M), 1.9 (M), 1.10 (M)	1.1, 1.8, 1.9, 1.10
Q2	2.1 (M), 2.2 (M), 2.3 (M), 2.9 (M), 2.13 (M)	2.1, 2.2, 2.3, 2.9, 2.13
Q3	3.1 (M), 3.4 (R)	3.1, 3.4
Q4	4.1 (M), 4.10 (M), 4.11 (M), 4.12 (M), 4.15 (M)	-
Q5	5.4 (M), 5.5 (M)	5.5
Q6	6.1 (M), 6.2 (M)	6.1, 6.2
Q7	-	-
Q8	8.1 (R)	8.1
Q9	-	-
Audio description		
Q1	1.1 (M), 1.3 (M), 1.4 (M), 1.6 (M), 1.8 (M)	1.1, 1.3, 1.4, 1.6
Q2	2.1 (M), 2.3 (M), 2.4 (R)	2.1, 2.3, 2.4
Q3	3.2 (M), 3.4 (R), 3.5 (R)	3.2, 3.4, 3.5
Q4	4.1 (M), 4.3 (R)	4.1, 4.3
Q5	-	-
Q6	6.1 (M), 6.3 (M), 6.6 (M), 6.7 (M), 6.8 (M), 6.12 (M)	6.1, 6.3, 6.6, 6.7, 6.12
Q7	7.1 (R), 7.2 (M)	7.1, 7.2
Q8	-	-
Q9	9.9 (R), 9.13 (M), 9.15 (M), 9.16 (R), 9.19 (R)	9.9, 9.13, 9.16
Transcript		
Q1	1.1 (M), 1.2 (M), 1.5 (R), 1.6 (R), 1.7 (R)	1.1, 1.2, 1.5, 1.6
Q2	2.1 (M), 2.2 (M)	2.1, 2.2
Q3	3.1 (M), 3.3 (M)	3.1, 3.3
Q4	4.1 (M), 4.2 (M)	4.1
Q5	5.1 (M)	5.1
Q6	6.1 (M)	6.1
Q7	-	-
Q8	-	-

Table 23 – Author 2, video 1 (documentary video)

Questions	Requirements applied	Requirements accomplished
Q1	1.1 (R), 1.8 (M), 1.9 (M), 1.10 (M)	1.1, 1.8, 1.9
Q2	2.1 (M), 2.2 (M), 2.3 (M), 2.9 (M), 2.13 (M)	2.1, 2.3, 2.9, 2.13
Q3	3.1 (M), 3.4 (R)	3.1
Q4	4.1 (M), 4.10 (M), 4.11 (M), 4.12 (M), 4.15 (M)	4.11, 4.12
Q5	5.4 (M), 5.5 (M)	5.5
Q6	6.1 (M), 6.2 (M)	6.2
Q7	-	-
Q8	8.1 (R)	-
Q9	-	-
Audio description		
Q1	1.1 (M), 1.3 (M), 1.4 (M), 1.6 (M), 1.8 (M)	1.1, 1.3, 1.6
Q2	2.1 (M), 2.3 (M), 2.4 (R)	2.1, 2.3
Q3	3.2 (M), 3.4 (R), 3.5 (R)	3.2, 3.5
Q4	4.1 (M), 4.3 (R)	4.1
Q5	-	-
Q6	6.1 (M), 6.3 (M), 6.6 (M), 6.7 (M), 6.8 (M)	6.3, 6.6, 6.8
Q7	7.1 (R), 7.2 (M)	7.1, 7.2
Q8	-	-
Q9	9.9 (R), 9.13 (M), 9.15 (M), 9.16 (R), 9.19 (R)	9.9, 9.13, 9.16
Transcript		
Q1	1.1 (M), 1.2 (M), 1.5 (R), 1.6 (R), 1.7 (R)	1.6, 1.7
Q2	2.1 (M), 2.2 (M)	2.1
Q3	3.1 (M), 3.3 (M)	3.1
Q4	4.1 (M), 4.2 (M)	4.1
Q5	5.1 (M)	5.1
Q6	6.1 (M)	6.1
Q7	-	-
Q8	-	-

Table 24 – Author 2, video 2 (animation video)

Questions	Requirements applied	Requirements accomplished
Q1	1.1 (R), 1.8 (M), 1.9 (M), 1.10 (M)	1.1, 1.8, 1.9, 1.10
Q2	2.1 (M), 2.2 (M), 2.3 (M), 2.6 (M), 2.9 (M), 2.13 (M)	2.1, 2.2, 2.3, 2.6, 2.9, 2.13
Q3	3.1 (M)	3.1
Q4	4.1 (M), 4.2 (R), 4.7 (M), 4.10 (M), 4.11 (M)	4.1, 4.2, 4.7, 4.10, 4.11
Q5	5.2 (M), 5.3 (M), 5.4 (M), 5.5 (M)	5.2, 5.3, 5.5
Q6	6.1 (M), 6.2 (M)	6.1, 6.2
Q7	7.2 (M)	7.2
Q8	8.3 (M)	8.3
Q9	-	-
Audio description		
Q1	1.1 (M), 1.4 (M), 1.6 (M), 1.8 (M)	1.1, 1.4, 1.6, 1.8
Q2	2.1 (M), 2.2 (M), 2.3 (M)	2.1, 2.2, 2.3
Q3	3.2 (M), 3.4 (R)	3.2, 3,4
Q4	4.1 (M), 4.3 (R)	4.1, 4.3
Q5	-	-
Q6	6.1 (M), 6.6 (M), 6.8 (M), 6.9 (M)	6.1, 6.6, 6.8, 6.9
Q7	7.1 (R), 7.2 (M), 7.3 (R)	7.1, 7.2
Q8	-	-
Q9	9.2 (M), 9.9 (R), 9.15 (M)	9.9, 9.15
Transcript		
Q1	1.1 (M), 1.2 (M), 1.5 (R), 1.6 (R), 1.7 (R)	1.1, 1.2, 1.5, 1.6
Q2	2.1 (M), 2.2 (M)	2.1, 2.2
Q3	3.1 (M), 3.3 (M)	3.1, 3.3
Q4	4.1 (M), 4.2 (M)	4.1
Q5	5.1 (M)	5.1
Q6	6.1 (M)	6.1
Q7	-	-
Q8	-	-

Table 25 – Author 3, video 1 (documentary video)

Questions	Requirements applied	Requirements accomplished
Q1	1.1 (R), 1.8 (M), 1.9 (M), 1.10 (M)	1.1, 1.8, 1.9, 1.10
Q2	2.1 (M), 2.2 (M), 2.3 (M), 2.9 (M), 2.13 (M)	2.1, 2.2, 2.3, 2.9, 2.13
Q3	3.1 (M), 3.4 (R)	3.1, 3.4
Q4	4.1 (M), 4.10 (M), 4.11 (M), 4.12 (M), 4.15 (M)	4.1, 4.11, 4.12, 4.15
Q5	5.4 (M), 5.5 (M)	5.5
Q6	6.1 (M), 6.2 (M)	6.1, 6.2
Q7	-	-
Q8	8.1 (R)	8.1
Q9	-	-
Audio description		
Q1	1.1 (M), 1.3 (M), 1.4 (M), 1.6 (M), 1.8 (M)	1.3, 1.4, 1.6, 1.8
Q2	2.1 (M), 2.3 (M), 2.4 (R)	2.3, 2.4
Q3	3.2 (M), 3.4 (R), 3.5 (R)	3.2, 3.4, 3.5
Q4	4.1 (M), 4.3 (R)	4.1, 4.3
Q5	-	-
Q6	6.1 (M), 6.3 (M), 6.6 (M), 6.7 (M), 6.8 (M)	6.1, 6.3, 6.6, 6.7
Q7	7.1 (R), 7.2 (M)	7.1, 7.2
Q8	-	-
Q9	9.9 (R), 9.13 (M), 9.15 (M), 9.16 (R), 9.19 (R)	9.9, 9.13, 9.15, 9.16, 9.19
Transcript		
Q1	1.1 (M), 1.2 (M), 1.5 (R), 1.6 (R), 1.7 (R)	1.1, 1.2, 1.5, 1.7
Q2	2.1 (M), 2.2 (M)	2.1, 2.2
Q3	3.1 (M), 3.3 (M)	3.1, 3.3
Q4	4.1 (M), 4.2 (M)	4.1, 4.2
Q5	5.1 (M)	5.1
Q6	6.1 (M)	6.1
Q7	-	-
Q8	-	-

Table 26 – Author 3, video 2 (animation video)

Questions	Requirements applied	Requirements accomplished
Q1	1.1 (R), 1.8 (M), 1.9 (M), 1.10 (M)	1.1, 1.8, 1.9, 1.10
Q2	2.1 (M), 2.2 (M), 2.3 (M), 2.6 (M), 2.9 (M), 2.13 (M)	2.1, 2.2, 2.3, 2.9, 2.13
Q3	3.1 (M)	3.1
Q4	4.1 (M), 4.2 (R), 4.7 (M), 4.10 (M), 4.11 (M)	4.1, 4.2, 4.7, 4.11
Q5	5.2 (M), 5.3 (M), 5.4 (M), 5.5 (M)	5.2, 5.3, 5.5
Q6	6.1 (M), 6.2 (M)	6.1, 6.2
Q7	7.2 (M)	7.2
Q8	8.3 (M)	8.3
Q9	-	-
Audio description		
Q1	1.1 (M), 1.4 (M), 1.6 (M), 1.8 (M)	1.4, 1.6, 1.8
Q2	2.1 (M), 2.2 (M), 2.3 (M)	2.2, 2.3
Q3	3.2 (M), 3.4 (R)	3.2, 3,4
Q4	4.1 (M), 4.3 (R)	4.1, 4.3
Q5	-	-
Q6	6.1 (M), 6.6 (M), 6.8 (M), 6.9 (M)	6.1, 6.6, 6.9
Q7	7.1 (R), 7.2 (M), 7.3 (R)	7.1, 7.2, 7,3
Q8	-	-
Q9	9.9 (R), 9.15 (M)	9.9, 9.15
Transcript		
Q1	1.1 (M), 1.2 (M), 1.5 (R), 1.6 (R), 1.7 (R)	1.1, 1.2, 1.5, 1.6, 1.7
Q2	2.1 (M), 2.2 (M)	2.1, 2.2
Q3	3.1 (M), 3.3 (M)	3.1, 3.3
Q4	4.1 (M), 4.2 (M)	4.1, 4.2
Q5	5.1 (M)	5.1
Q6	6.1 (M)	6.1
Q7	-	-
Q8	-	-

Table 27 – Author 4, video 1 (animation video)

Questions	Requirements applied	Requirements accomplished
Q1	1.1 (R), 1.8 (M), 1.9 (M), 1.10 (M)	1.1, 1.8, 1.9, 1.10
Q2	2.1 (M), 2.2 (M), 2.3 (M), 2.6 (M), 2.9 (M), 2.13 (M)	2.1, 2.2, 2.3, 2.13
Q3	3.1 (M)	3.1
Q4	4.1 (M), 4.2 (R), 4.7 (M), 4.10 (M), 4.11 (M)	4.1, 4.2, 4.10, 4.11
Q5	5.2 (M), 5.3 (M), 5.4 (M), 5.5 (M)	5.2, 5.3, 5.5
Q6	6.1 (M), 6.2 (M)	6.2
Q7	7.2 (M)	7.2
Q8	8.3 (M)	8.3
Q9	-	-
Audio description		
Q1	1.1 (M), 1.4 (M), 1.6 (M), 1.8 (M)	1.1, 1.6, 1.8
Q2	2.1 (M), 2.2 (M), 2.3 (M)	2.1, 2.2, 2.3
Q3	3.2 (M), 3.4 (R)	3.2
Q4	4.1 (M), 4.3 (R)	4.1
Q5	-	-
Q6	6.1 (M), 6.6 (M), 6.8 (M), 6.9 (M)	6.1, 6.6, 6.9
Q7	7.1 (R), 7.2 (M), 7.3 (R)	7.1, 7.2, 7,3
Q8	-	-
Q9	9.2 (M), 9.9 (R), 9.15 (M), 9.26 (R)	9.9, 9.15, 9.26
Transcript		
Q1	1.1 (M), 1.2 (M), 1.5 (R), 1.6 (R), 1.7 (R)	.6, 1.7
Q2	2.1 (M), 2.2 (M)	2.1, 2.2
Q3	3.1 (M), 3.3 (M)	3.1, 3.3
Q4	4.1 (M), 4.2 (M)	4.1, 4.2
Q5	5.1 (M)	5.1
Q6	6.1 (M)	6.1
Q7	-	-
Q8	-	-

Table 28 – Author 4, video 2 (documentary video)

Questions	Requirements applied	Requirements accomplished
Q1	1.1 (R), 1.8 (M), 1.9 (M), 1.10 (M)	1.1, 1.8, 1.9, 1.10
Q2	2.1 (M), 2.2 (M), 2.3 (M), 2.9 (M), 2.13 (M)	2.1, 2.2, 2.3, 2.13
Q3	3.1 (M), 3.4 (R)	3.1, 3.4
Q4	4.1 (M), 4.10 (M), 4.11 (M), 4.12 (M), 4.15 (M)	4.1, 4.10, 4.11, 4.12, 4.15
Q5	5.4 (M), 5.5 (M)	5.5
Q6	6.1 (M), 6.2 (M)	6.1, 6.2
Q7	-	-
Q8	8.1 (R)	8.1
Q9	-	-
Audio description		
Q1	1.1 (M), 1.3 (M), 1.4 (M), 1.6 (M), 1.8 (M)	1.1, 1.3, 1.4, 1.6
Q2	2.1 (M), 2.3 (M), 2.4 (R)	2.1, 2.4
Q3	3.2 (M), 3.4 (R), 3.5 (R)	3.2, 3.4, 3.5
Q4	4.1 (M), 4.3 (R)	4.1
Q5	-	-
Q6	6.1 (M), 6.3 (M), 6.6 (M), 6.7 (M), 6.8 (M)	6.1, 6.3, 6.6, 6.7
Q7	7.1 (R), 7.2 (M)	7.1, 7.2
Q8	-	-
Q9	9.9 (R), 9.13 (M), 9.15 (M), 9.16 (R), 9.19 (R)	9.9, 9.13, 9.15, 9.16
Transcript		
Q1	1.1 (M), 1.2 (M), 1.5 (R), 1.6 (R), 1.7 (R)	1.1, 1.2, 1.5, 1.6, 1.7
Q2	2.1 (M), 2.2 (M)	2.1, 2.2
Q3	3.1 (M), 3.3 (M)	3.1, 3.3
Q4	4.1 (M), 4.2 (M)	4.1, 4.2
Q5	5.1 (M)	5.1
Q6	6.1 (M)	6.1
Q7	-	-
Q8	-	-

TRANSCRIPT TEMPLATE

Figure 23 – Transcript template

Type of video content: Title, Month, Year

Abstract. An abstract is a short explication of video content, it also could contain where the video was recorded, and who are the main actors (in case of presentations, who is presenting).

- [Video with audio and slides. Title](#)
- [Downloadable version of video with audio \(mp4\)](#)
- [Downloadable version of slides \(.ppt\)](#)
- Text transcript is below

Transcript

Johana: Good morning. Today, I going to explain some guidelines for transcript.

[Guidelines about speakers names]

First, it is important include the speakers names. Capitalize proper nouns and use colon at the end, for example, "John: Hello". The W3C in their transcripts uses bold for names, for example, "**John:** Hello". For an informal podcast transcript with multiple speakers, it's probably best to use speakers' full names the first time, and then just their first names. When there are multiple speakers, use hanging indents to make it easy to skim for a particular speaker. It is possible use CSS and give each speaker a different color or style. (Make sure to include names so there is redundant color coding.) When the focus should be on the interviewee's answers and not the interviewer, format the interviewers questions in smaller, lighter text so the interviewee's answers stand out more clearly.

[Guidelines about relevant visual information]

Also include relevant visual information, that means include visual context, actions and expressions of actors, and any other relevant visual information (photografies, unusual dress). All additional information should be in brackets. For example "[photografy of a woman with a blue macaw inside a picture frame that says best friends]", "Joe: I hate this computer! [shouted]".

[Guidelines about navigation and clarifications]

You could add navigation and clarifications. Add headings and links where it will make the transcript more usable. For example, add links when the video is short and explain a lot of concepts, create links for these concepts; or add titles when the video is long. Add clarifying information, as long as it is clear that it is not part of the actual audio, e.g., words added to a paragraph put in [brackets], or separate sections with headings "Introduction", "Transcript", "Resources". To know more guidelines read the Chapter 4 of this thesis. Thank you.

[Clapping]

Audience: [inaudible]

TRANSCRIPT4ALL TOOL

Transcript4All is an authoring tool for transcript, which provide some functionalities according the transcript template created in this research (Appendix C) and following the accessibility recommendations (Section 3.4.3). The Transcript4All is available on the link cafe.intermidia.icmc.usp.br:22080/transcript4all.

The Transcript4All is available in three languages: Portuguese English and Spanish. It provide the help functionality with shortcuts. The tool is composed by four activities: creating the project, obtaining the video information, creating the transcript and exportation.

The first activity, **creating the project**, consist in the project creation from a YouTube's video link or from a local video. If the author has the caption for the video, the Transcript4All provide the option to begins the transcription using these caption, otherwise the author begins from zero.

The second activity, **editing the video information**, consist in include the type of video content, title, data, duration, and abstract. Type of video would be presentation, podcast, documentary, cooking recipes, etc. So that the final format would be "Presentation: Video accessibility, February 8th 2016." An abstract is a short explication of video content, it also could contain where the video was recorded, and who are the main actors (in case of presentations, who is presenting). Figure 24 shows the interface for this activity.

The third activity, **creating the transcript**, consist in describe the video and the Transcript4All provide some functionalities such as: provide an accessible video player to help the author with the transcription edition, provide a set of recommendations as a guide to amateur authors and provide an edition area. Figure 25 shows the interface for this activity. The recommendations are based on Lawton (2009) and DigitalGov (2013) and their aim is help amateur authors to create transcriptions. Each recommendation is composed by a description, an example and a figure representing that example. The recommendations are:

Figure 24 – Transcript4All - Video information

The screenshot displays the 'Transcript4All' web interface. At the top left is the 'Video4All' logo, and at the top right are 'Help' and 'English' buttons. The main heading is 'Transcript authoring'. On the left side, there is a video player area with a play button and a transcript box. The transcript box contains the text: 'Julia: Hi Pedro, how are you?'. Below the transcript box is a video player with a play button and a progress bar. On the right side, there is a 'Video information' form with the following fields: Title (text input), Category (dropdown menu), Abstract (text area), Record date (text input), Publication date (text input), and Duration (text input). A green 'Next' button is located at the bottom of the form.

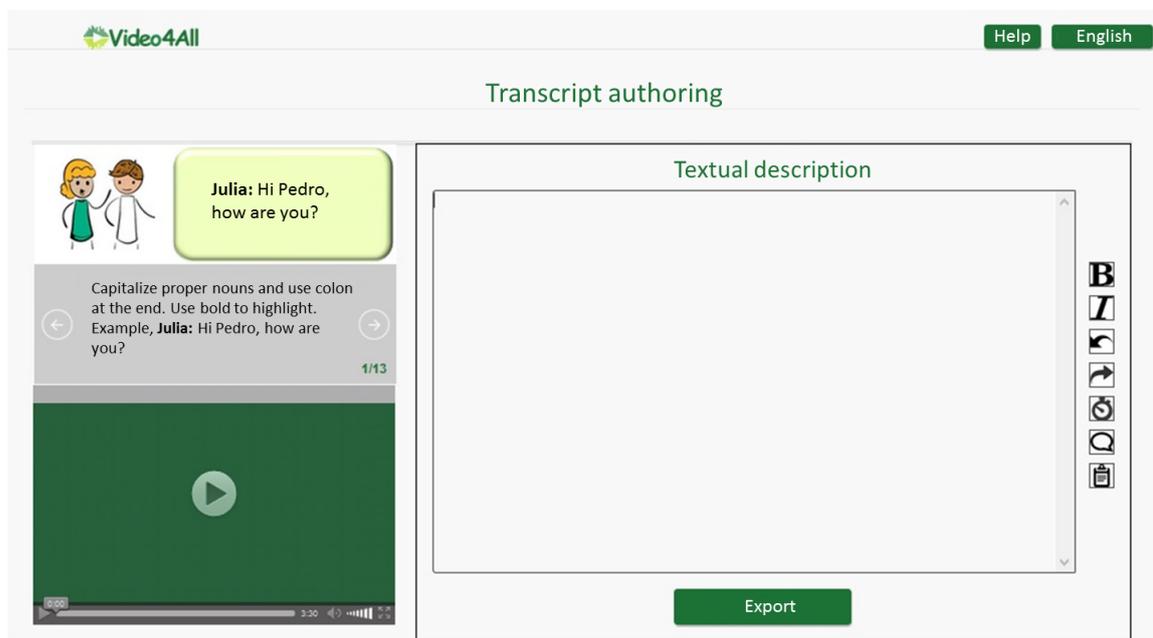
Source: Elaborated by the author.

1. Capitalize proper nouns and use colon at the end. Use bold to highlight. For example, **Julia:** Hi Pedro, how are you?.
2. Separate paragraphs by actors. Each paragraph should contain all speech of that actor.
3. Include all speech content. If there is speech that is not relevant, indicate that it has been excluded from the transcript. For example, [participants discuss the weather while the presenter reboots his computer].
4. Include relevant information. Use brackets, for example, Joe: I hate this computer! [shouted].
5. Include relevant visual information (actions and expressions of actors, photographs, unusual dress). For example, [photography of a woman with a blue macaw inside a picture frame that says best friends].
6. Include relevant non-speech audio (use of irony, music and scenery). The irrelevant background noise is generally not described. For example, [rose petals are thrown from a helicopter].
7. Edit according to situation. In some cases, such as legal depositions, the transcript must be verbatim, including ums, ahs, and indicating pauses. For example, **Ana:** umm! [remains silent].
8. When there are multiple speakers, it's probably best to use speakers' full names the first time, and then just their first names.

9. For most podcasts, presentations, and such, minor edits for readability are appropriate. For example, it would generally be correct to edit “The score of the game was 1 - 0. I mean 2 - 0” to “The score of the game was 2 - 0.”
10. It is not appropriate to change the meaning from the audio. Likewise, it is usually not appropriate to significantly correct grammar or other mistakes.
11. Add titles when the video is long. It helps with the Search Engine Optimization.
12. Add links where it will make the transcript more usable. For example, when the video is short and explains many concepts, create links for these concepts.
13. Add clarifying information, as long as it is clear that it is not part of the actual audio, e.g., words added to a paragraph put in [brackets], or separate sections with headings. For example, [Introduction], [Concepts], [Conclusion].

The edition area provides functionalities to help authors applying the recommendations without understand HTML code. These functionalities are: setting the text in bold, adding a character with the format recommended (capitalize proper nouns and colon at the end), adding links and adding titles.

Figure 25 – Transcript4All - Textual description



Source: Elaborated by the author.

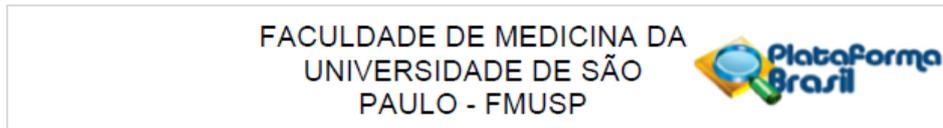
Finally, the fourth activity is **exportation**, it consists in export the transcript in two formats, textual and HTML. For this activity it was used the transcript template (Appendix C).

REPORT OF ETHIC COMMITTEE

Figure 26 – Report of Ethic Committee

<p>FACULDADE DE MEDICINA DA UNIVERSIDADE DE SÃO PAULO - FMUSP</p> 
PARECER CONSUBSTANCIADO DO CEP
DADOS DO PROJETO DE PESQUISA
<p>Título da Pesquisa: Interação com conteúdo multimídia para apoio à geração de vídeos acessíveis Pesquisador: Johana Maria Rosas Villena Área Temática: Versão: 1 CAAE: 45075015.8.0000.0065 Instituição Proponente: UNIVERSIDADE DE SAO PAULO Patrocinador Principal: Financiamento Próprio</p>
DADOS DO PARECER
<p>Número do Parecer: 1.073.606 Data da Relatoria: 20/05/2015</p>
<p>Apresentação do Projeto: Trata-se de projeto de doutorado pelo Instituto de Ciências Matemáticas e da Computação de ICMC/USP/SC, encaminhado para análise ao CEP da EACH/USP. O projeto de pesquisa visa investigar a interação de sujeitos com conteúdo multimídia para apoio à geração de vídeos acessíveis. Será explorada incorporação pelo sujeitos de elementos para tornar o vídeo acessível, como legendas, transcrição, audiodescrição e linguagem de sinais. Alunos e professores do ICMC serão convidados a participar de um experimento, estima-se a participação de 4 sujeitos. A seleção dos participantes será feita por conveniência por meio de um convite que será enviado por e-mail. Dar-se-á preferência a aqueles com algum tipo de deficiência, os mais velhos, os menos escolarizados e os de classe social mais baixo e que tenham menos contato com o computador. Considera que pessoas com esses perfis poderão utilizar em maior escala as características de acessibilidade geradas pelo método. Se houver muitos interessados, considera-se a possibilidade de abrir um curso de extensão. O experimento será precedido e sucedido por preenchimento de um questionário. Ainda antes do experimento os sujeitos participação de treinamento para conhecer e gerar recursos de acessibilidade. O experimento compreenderá realização pelos sujeitos individualmente de tarefas para gerar elementos para tornar o vídeo acessível. Os participantes receberão material de apoio e explicações sobre o método e, treinamento adequado para cada uma das tarefas. O questionário inicial compreende alguns dados e características</p>
<p>Endereço: DOUTOR ARNALDO 251 21º andar sala 36 Bairro: PACAEMBU CEP: 01.246-903 UF: SP Município: SAO PAULO Telefone: (11)3893-4401 E-mail: cep_fm@usp.br</p>

Figure 27 – Report of Ethic Committee



Continuação do Parecer: 1.073.606

personais. O experimento compreenderá no máximo 3 horas, incluindo o Questionário Pós-Sessão, relacionado a opinião do sujeito sobre o experimento. interação .

Objetivo da Pesquisa:

O foco primário é a porcentagem de diretrizes atendidas por cada elemento. O foco secundário é a quantidade de requisitos de acessibilidade atendidos. O foco qualitativo primário é avaliar o método para geração de vídeos acessíveis coletando os benefícios durante a criação de elementos. O foco secundário é avaliar as dificuldades durante a utilização do método.

Avaliação dos Riscos e Benefícios:

Prevê-se riscos e como contorná-lo. Para amenizar o risco a exposição à radiação, os participantes não ficarão mais do que 4 (quatro) horas em frente ao computador (tempo considerado desprezível). Para amenizar o risco de violação da privacidade, os dados coletados não estarão expostos na Web. Para amenizar o risco relacionado ao tempo despendido, bem como o tempo gasto para responder os questionários, o conteúdo dos vídeos e dos questionários serão escolhidos e planejados de modo a evitar possíveis constrangimentos ou desconfortos. Os benefícios são considerados diretos e relacionados à geração de vídeos acessíveis e, indiretos relacionados a colaboração com o avanço das pesquisas sobre o tema. Os sujeitos da pesquisa podem colaborar ainda com sugestões de melhorias no método de geração dos elementos ou mesmo no reprodutor de vídeos acessível.

Comentários e Considerações sobre a Pesquisa:

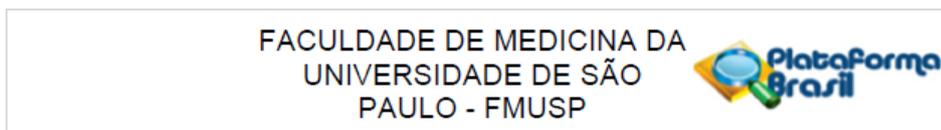
Projeto bastante sucinto quanto à delimitação e justificativa do estudo e descrição do experimento, mas suficientemente claro para sua compreensão. Embora os critérios de inclusão da população no estudo não tenham implicações éticas desfavoráveis, considero que para maior aproveitamento da interação deveria se assegurar inclusão de pessoas com algum tipo de deficiência sensorial no estudo, por usarem em maior escala as características de acessibilidade geradas pelo método.

Considerações sobre os Termos de apresentação obrigatória:

Constam todos os documentos necessários à submissão estando o TCLE redigido de modo acessível e cuidadoso quanto aos aspectos éticos

Endereço: DOUTOR ARNALDO 251 21º andar sala 36
 Bairro: PACAEMBU CEP: 01.246-903
 UF: SP Município: SAO PAULO
 Telefone: (11)3893-4401 E-mail: cep.fm@usp.br

Figure 28 – Report of Ethic Committee



Continuação do Parecer: 1.073.606

Recomendações:

assegurar inclusão de pessoas com algum tipo de deficiência sensorial no estudo visando sua potencial maior interação e incorporação nos vídeos de elementos para torná-los acessíveis.

Conclusões ou Pendências e Lista de Inadequações:

sem pendências quanto aos aspectos éticos

Situação do Parecer:

Aprovado

Necessita Apreciação da CONEP:

Não

Considerações Finais a critério do CEP:

SAO PAULO, 21 de Maio de 2015

Assinado por:
Roger Chammas
(Coordenador)

Endereço: DOUTOR ARNALDO 251 21º andar sala 36
Bairro: PACAEMBU **CEP:** 01.246-903
UF: SP **Município:** SAO PAULO
Telefone: (11)3893-4401 **E-mail:** cep.fm@usp.br

GUIDELINES FOR CLOSED CAPTION

The Captioning Key contains guidelines for caption, published by the The Described and Captioned Media Program (DCMP) on the web page <<http://www.captioningkey.org/>>. These guidelines are available to the public and they are consistent with the 2014 mandates by the Federal Communications Commission (FCC) of United States.

B.1 Text

Text refers to the appearance and presentation of the letters and words. Text considerations include case, font, line division, and caption placement.

Case

Mixed case characters are preferred for readability. However, capital letters are used for screaming or shouting.

Font

A font, or typeface, is a set of characters of a certain size, weight, and style. Font characteristics must be consistent throughout the media.

1. The use of white characters is preferred. They need to be medium weight, be sans serif, have a drop or rim shadow, and be proportionally spaced. The use of a translucent box is preferred so that the text will be clearer, especially on light backgrounds.
2. The font must include upper- and lowercase letters with descenders that drop below the baseline.

Line Division

When a sentence is broken into two or more lines of captions, it should be broken at a logical point where speech normally pauses.

1. Do not break a modifier from the word it modifies. Example:

Inappropriate

Mark pushed his black
truck.

Appropriate

Mark pushed
his black truck.

2. Do not break a prepositional phrase. Example:

Inappropriate

Mary scampered under
the table.

Appropriate

Mary scampered
under the table.

3. Do not break a person's name nor a title from the name with which it is associated.
Example:

Inappropriate

Bob and Susan
Smythe are at the movies.

Appropriate

Bob and Susan Smythe
are at the movies.

4. Do not break a line after a conjunction. Example:

Inappropriate

In seconds she arrived, and
he ordered a drink.

Appropriate

In seconds she arrived,
and he ordered a drink.

5. Do not break an auxiliary verb from the word it modifies. Example:

Inappropriate

Mom said I could
have gone to the movies.

Appropriate

Mom said I could have gone
to the movies.

6. Never end a sentence and begin a new sentence on the same line unless they are short, related sentences containing one or two words. Example:

Inappropriate

He suspected that his face
turned pale. He knew he
wouldn't be able to speak
if spoken to. Running toward
the void, he halted. . .

Appropriate

He suspected that his face
turned pale.
He knew he wouldn't be able
to speak if spoken to.
Running toward the void,
he halted. . .

Caption Duration

Captions for broadcast require a build-up time, also known as load time, in order for the captions to display properly. The build-up time is primarily dependent on the amount of text in the caption as calculated from the in-point of the current caption to the in-point of the subsequent caption. Captions prepared for broadcast or the DCMP must not include build-time errors.

1. Captions should have a minimum duration of 40 frames (1 second and 10 frames).

2. Captions should have a maximum duration of 6 seconds. (Background music notation is an exception to this guideline).

Caption Placement

Caption placement (vertical and horizontal) refers to the location of captions on the screen. Placement is not possible on all players, but when it is, the following guidelines are important for clarity.

1. Multi-lined captions should be left aligned when technically possible.
2. Captions are placed on the bottom two lines, but placement should not interfere with names, faces, or mouths of speakers or text/graphics that are essential to the comprehension of the media. Should interference occur, captions should be placed at the top of the screen. If placing captions at the top of the screen also interferes with visuals/graphics, place captions elsewhere on the screen where they do not interfere.
3. It is preferred that there are no more than two lines per caption.
4. It is essential to place all captions within reasonable margins. This will avoid the possibility of missing characters at right or left screen or missing descenders/ascenders at bottom or top screen.
5. Captioned dialogue must be placed under the speaker as long as it does not interfere with graphics or other preexisting features. For media with one offscreen narrator and no preexisting graphics, captions should be left-aligned at center screen on the bottom two lines.
6. Three- or four-line captions are occasionally acceptable if a one- or two-line caption would interfere with preexisting graphics or be confusing with regard to speaker identification. Video still of two girls sitting on a lake dock. Captions read: "Oooh, it's gorgeous!" and "It's so beautiful!".
7. When people onscreen speak simultaneously, place the captions underneath the speakers. If this is not possible due to the length of the caption or interference with onscreen graphics, caption each speaker at different timecodes. Do not use other speaker identification techniques, such as hyphens.
8. If a speaker continuously moves from one onscreen location to another, one placement for captions of that speaker's communication must be used. Speaker identification may be added for clarification. (Confusion occurs when captions jump around the screen.)

B.2 Language Mechanics

Language mechanics incorporate the proper use of spelling, capitalization, punctuation, grammar, and other factors deemed necessary for high-quality captioned media. Rules included in these guidelines are primarily those which are unique to captioning and speech-to-text. For conventional words, dictionaries and style guides must be followed. Proper names, technical terms, and specialized language must be verified through specialty references or directly from an authoritative source.

Spelling and Capitalization

1. Do not emphasize a word using all capital letters except to indicate screaming or shouting.
2. Be consistent in the spelling of words throughout the media. This includes vocabulary that can be spelled either as one or two words or in hyphenated form.
3. Capitalize proper nouns for speaker identification. All other speaker identification should be lowercased unless this identification is being used as a proper noun.
4. Lowercase sound effects, including both description and onomatopoeia, except when a proper noun is part of the description.
5. See the Numbers section on the Captioning Key Appendices page for detailed guidelines for numbers including dates, time, periods of time, fractions, percents, dollar amounts, and measurements.

Punctuation and Grammar

Commas

1. When captioning a list separated by commas, use a serial, or Oxford, comma. Example:

Inappropriate

I'm having eggs, bacon and toast.

Appropriate

I'm having eggs, bacon, and toast.

Hyphens and Dashes

1. When a speaker hesitates or stutters, caption what is said. Correct Example: c-c-c-old
2. When captioning spelling (including fingerspelling), separate capital letters with hyphens. Correct Example: A-N-T-I-O-N-E-T-T-E

Ellipses

1. Use an ellipsis when there is a significant pause within a caption.
2. Do not use an ellipsis to indicate that the sentence continues into the next caption.
3. Use an ellipsis to lead into or out of audio relating to an onscreen graphic. Video still says 130 million of the world's youth are not attending school. Caption reads '...70% are girls.'

Quotation Marks

1. Use quotation marks for onscreen readings from a poem, book, play, journal, or letter. However, use quotation marks and italics for offscreen readings or voice-overs.
2. Beginning quotation marks should be used for each caption of quoted material except for the last caption. The last caption should have only the ending quotation mark.

Correct Example

Reading from a journal. . .

"Mother knelt down
and began thoughtfully fitting
"the ragged edges
of paper together.

The process was watched
with spellbound interest."

Spacing

1. Spaces should not be inserted before ending punctuation, after opening and before closing parentheses and brackets, before and after double hyphens and dashes, or before/between/after the periods of an ellipsis.
2. A space should be inserted after the beginning music icon (§) and before the ending music icon(s). Correct Example: § There's a bad moon rising §

Italics

Use italics as follows:

1. A voice-over reading of a poem, book, play, journal, letter, etc. (This is also quoted material, so quotation marks are also needed.)

2. When a person is dreaming, thinking, or reminiscing.
3. When there is background audio that is essential to the plot, such as a PA system or TV.
4. The first time a new word is being defined, but do not italicize the word thereafter.
5. Offscreen dialogue, narrator (see Exception 2 below), sound effects, or music (this includes background music).
6. The offscreen narrator when there are multiple speakers onscreen or offscreen.
7. Speaker identification when the captioned dialogue is in italics.
8. Foreign words and phrases, unless they are in an English dictionary.
9. When a particular word is heavily emphasized in speech. Correct Example: You *must* go!

Exceptions to the use of italics include:

1. When an entire caption is already in italicized format, use Roman type to set off a word that would normally be italicized.
2. If there is only one person speaking throughout the program (including the narrator), whether onscreen or offscreen, use Roman type with no italics.
3. Do not italicize when a person who is offscreen is translating for a speaker who is onscreen.

B.3 Presentation Rate

A Definition of Presentation Rate

The presentation rate is the number of captioned words per minute (wpm) that are displayed onscreen. A reoccurring question about captioning is whether captions should be verbatim or edited. Among the advocates for verbatim are organizations of deaf and hard of hearing persons who do not believe that their right for equal access to information and dialogue is served by any deletion or change of words. Supporters of edited captions include parents and teachers who call for the editing of captions on the grounds that the reading rates necessitated by verbatim captions can be so high that captions are almost impossible to follow.

As the debate has continued, researchers have tackled the question. A bibliography of research on reading rates is provided in the Captioning Presentation Rate Research document on the Captioning Key Appendices page. DCMP supports editing based on research results and the DCMP's half-century of captioning experience. Editing is often essential to ensure that students have time to read the captions, integrate the captions and picture, and internalize and comprehend the message.

When editing occurs, each caption should maintain the meaning, content, and essential vocabulary of the original narration. DCMP media users, who are the families and teachers of students who are deaf and hard of hearing, have enthusiastically praised the quality of the DCMP educational media and the captioning that provides equal access.

Specifications

Each word is counted when calculating the presentation rate, as opposed to basing the calculation on the number of characters. In addition, speaker identification, sound effects, and other similar elements must be included in the calculations.

All lower-level educational media should optimally be captioned at a presentation rate not to exceed 130 words per minute (wpm), middle-level not to exceed 140 wpm, and upper-level not to exceed 160 wpm. However, as stated above, the original meaning, content, and essential vocabulary must be maintained. Achieving optimum rates is more difficult when complex new concepts and vocabulary are introduced in a media production.

The only times when presentation rate is ignored are when any person is quoted, a well-known or famous person is speaking on-screen or off-screen, poems and other published works are quoted, and/or song lyrics are sung. These must be captioned verbatim.

Guidelines

One way of achieving the desired presentation rate is by borrowing 15 frames before or after the audio occurs. This "borrowing" technique is hardly noticeable to the viewer.

An alternative way to achieve the rate is by editing. Following are two techniques for editing and examples:

1. Editing can be relatively simple by eliminating redundant and/or nonessential information. Following are examples of this approach:

Original Narration

Will you get out of here!

Edited

Will you get out!

Original Narration

It is really, really difficult to find good help.

Edited

It is really difficult

to find good help.

Original Narration

That, uh, color–um, this shade of blue matches your eyes.

Edited

This shade of blue matches your eyes.

2. Editing becomes more complex when it is necessary to move beyond the simple elimination of non-essential information and may involve shortening, deleting, or reordering words, phrases, or sentences. Such editing must maintain the original meaning, content, and essential vocabulary. Following are examples of this approach:

Original Narration

I found that it was a lot easier to keep track of the words and the lines.

Edited

I found it much easier to keep
track of words and lines.

Original Narration

And even though they're looking at the word, their focus is going elsewhere.

Edited

They're looking at the word,
but their focus is elsewhere.

Original Narration

And what we're saying is that people with dyslexia tend to get distracted by the words on either side.

Edited

Words on either side tend to
distract people with dyslexia.

Original Narration

They rely not only on the organic matter
that's being produced,
but also the complexity of the habitat
to hide and escape from predators.

Edited

They rely on organic matter that's being produced
and habitat complexity to hide and escape from predators.

B.4 Sound Effects and Music

Sound Effects

Sound effects are sounds other than music, narration, or dialogue. They are captioned if it is necessary for the understanding and/or enjoyment of the media.

1. A description of sound effects, in brackets, should include the source of the sound. However, the source may be omitted if it can be clearly seen onscreen. Examples: The video shot shows a football game with man running with the ball. Caption reads: [audience cheering]. The video shot shows a thick smoke. Caption reads: [explosions].
2. Described sound effects may be combined with onomatopoeia. The described sound effect should be on the first line of the caption, separate from the onomatopoeia. Both described sound effects and onomatopoeias must be lowercased. Example: The video shot shows a fisherman on a boat. Caption reads: [engine idling] rrrrr.
3. Offscreen sound effects should be italicized, if italics are available.
4. Place the description of the sound effect as close as possible to the sound source.
5. For offscreen sound effects, it is not necessary to repeat the source of the sound if it is making the same sound a few captions later.

Correct Example

First Caption

[pig squealing]

Second Caption

[squealing continues]

6. Use punctuation to indicate speed or pace of sound.

Correct Example

Slow

[clock chiming] dong...dong...dong

Rapid

[gun firing] bang, bang, bang

7. A sound represented by a repeated word is not hyphenated. A sound represented by two different words is hyphenated.

Correct Example

Repeated Words

[doorbell ringing] ding, ding

Two Different Words

[doorbell ringing] ding-dong

8. When describing a sustained sound, use the present participle form of the verb. When describing an abrupt sound, use the third person verb form.

Correct Example

Sustained Sound

[dog barking] woof, woof. . . woof

Abrupt Sound

[dog barks] woof!

9. Caption background sound effects only when they are essential to the plot.
10. When possible, use specific rather than vague, general terms to describe sounds.

Correct Example

Vague/General

[bird singing]

Specific

[robin singing]

11. Never use the past tense when describing sounds. Captions should be synchronized with the sound and are therefore in the present tense.

Music

Background Music

1. A description (in brackets) should be used for instrumental/background music when it's essential to the understanding of the program.
2. Offscreen background music description should be italicized.
3. If possible, the description should include the performer/composer and the title.
4. Use descriptions that indicate the mood. Be as objective as possible. Avoid subjective words, such as "delightful," "beautiful," or "melodic."

Correct Examples

[Louis Armstrong plays "Hello Dolly"]

[lyrical flute solo]

[pianist playing the national anthem]

5. Beware of misplaced modifiers in descriptions.

Incorrect

[frantic piano playing]

Correct

[frantic piano music]

6. Nonessential background music should be captioned by placing a music icon (§) in the upper right corner of the screen and should never be captioned at the expense of dialogue.
7. Do not caption background music with a duration under 5 seconds.

Lyrics

1. If music contains lyrics, caption the lyrics verbatim. The lyrics should be introduced with the name of the artist and the title in brackets, if the presentation rate permits. Example: the video shot shows clouds over the ocean. Caption reads: [Dinah Washington singing "Mad about the Boy"].
2. Caption lyrics with music icons (§). Use one music icon at the beginning and end of each caption within a song, but use two music icons at the end of the last line of a song. A space should be inserted after the beginning music icon (§) and before the ending music icon(s). Example: the video shot shows a dark screen. Caption says: (Musical note) The boys are back in town" (musical notes).

B.5 Speaker Identification

Establishing the identity of both onscreen and offscreen speakers is vital for clarity.

1. When possible, use caption placement to identify an onscreen speaker by placing the caption under the speaker.
2. Do not identify the speaker by name until the speaker is introduced in the audio or by an onscreen text/graphic.
3. When a speaker cannot be identified by placement and his/her name is known, the speaker's name should be in parentheses.

The speaker's name needs to be on a line of its own, separate from the captions.

Correct Example

(Jack)

I don't see how blasting
would work on this building.

4. When a speaker cannot be identified by placement and his/her name is unknown, identify the speaker using the same information a hearing viewer has.

Correct Examples

(female N^o1)

I'm happy.

(male narrator)

That is wonderful!

5. If a speaker is offscreen and his/her offscreen position is known, place the captions to the far right or left, as close as possible onscreen to the offscreen speaker's assumed position.
6. If there are multiple speakers and only one narrator, identify as (female narrator) or (male narrator) at the beginning of the media. It is not necessary to identify gender for each caption thereafter.
7. When an actor is portraying another person or character, identify the actor as the person being portrayed.

Correct Example

(as George Washington)

If the freedom of speech
is taken away,
then dumb and silent
we may be led,
like sheep
to the slaughter.

B.6 Special Considerations

Spoken language is rich and full of meaning. However, it also consists of oddly formed sentences and even word play. Accuracy, clarity, and readability are challenges.

Intonation, Play on Words, and No Audio

1. If the speaker is not visible onscreen, or visual clues that denote the emotional state are not shown, indicate the speaker's emotion.

Correct Example

[angrily]

Well, whatever!

2. When a person is whispering, caption as:

Correct Example

[whispering]

Okay, you go first.

3. When feasible, describe puns.

Correct Example

Why do they call her "Ouisy"? ["Wheezy"]

4. When people are seen talking, but there is no audio, caption as [no audio] or [silence].

Foreign Language, Dialect, Slang, and Phonetics

1. If possible, caption the actual foreign words. If it is not possible to caption the words, use a description (e.g., [speaking French]). Never translate into English.
2. If possible, use accent marks, diacritical marks, and other indicators.
3. Indicate regional accent at the beginning of the first caption.

Correct Example

[Southern accent]

If y'all want me to.

4. Keep the flavor of dialect.

Correct Example

You sho' ain't

from 'round here.

5. Caption profanity and slang if in the audio.

6. When a word is spoken phonetically, caption it the way it is commonly written.

Original Narration

"N-double-A-C-P"

Captioned As

NAACP

Original Narration

"www dot D-C-M-P dot org"

Captioned As

www.dcmp.org

Original Narration

"eight or nine hundred"

Captioned As

800 or 900

Original Narration

"a thousand"

Captioned As

a thousand

Original Narration

"One thousand"

Captioned As

1000

B.7 Appendices

Captioning Presentation Rate Research

This appendix is a research document which contains a conglomerate of studies related to both children and adults and how they view, read, and prefer captions. Initially, it seems to be common sense that verbatim captioning is the ideal, the mark of true equal access. However, it may be possible for spoken audio to be delivered so quickly that most people cannot read its verbatim captioning, which seems counter-productive to the goal of equal access.

Additional research is needed to determine how fast captions should appear on the screen and what presentation rates people prefer. These critically important issues have only partially

been addressed. But there is an existing body of research and study that supports the DCMP policy and philosophy. More information on the website¹.

Numbers

Experts don't always agree on rules for writing numbers or numerals. Captionists should follow a standard style manual, remembering to be consistent, or use this relatively detailed overview.

Spelling Out

1. Unless otherwise specified below, spell out all numbers from one to ten, but use numerals for all numbers over ten. Examples:

Inappropriate The fifty-four DVDs
need to be shelved.

Appropriate

The 54 DVDs
need to be shelved.

Inappropriate

He's at the thirty, the twenty,
and scores!

Appropriate

He's at the 30, the 20,
and scores!

2. Spell out any number that begins a sentence as well as any related numbers. Example:

Two hundred guests
and eleven guides entered.

3. Spell out casual, nonemphatic numbers. Example:

He gave me
hundreds of reasons.

4. Numerals with four digits can either have a comma or not. Be consistent throughout the media production. For numerals having over five digits, a comma is necessary. Example:

Inappropriate

¹ <http://www.captioningkey.org/appendices.html>

50000

Appropriate

50,000

5. Use numerals in a listing of numbers if one or more is above ten and these occur in one caption or one sentence. Example:

Inappropriate

Steven has 21 books,

11 oranges, and three cats.

Appropriate

Steven has 21 books,

11 oranges, and 3 cats.

6. Use numerals when referring to technical and athletic terms. Example:

He scored 3 goals

in today's game!

7. When indicating sequence, capitalize the noun and use numerals. Exceptions are the indication of line, note, page, paragraph, size, step, or verse. Examples: Building 2, page 31, Channel 5, size 12, Chapter III, step 3, Room 438, paragraph 2.

Dates

1. Use the numeral plus the lowercase "th," "st," or "nd" when a day of the month is mentioned by itself (no month is referred to). Example:

Captioned As

Bob went fishing

on the 9th.

Original Narration

"ninth"

2. When the day precedes the month, use the numeral plus the lowercase "th," "st," or "nd" if the ending is spoken. Example:

Captioned As

My birthday is

the 17th of June.

Original Narration

"seventeenth"

3. Use the numeral alone when the day follows the month. Example:

Captioned As

I will meet you
on May 9.

Original Narration

"nine" or "ninth"

4. When the month, day, and year are spoken, use the numeral alone for the day, even if an ending ("th," "st," or "nd") is spoken. Example:

Captioned As

Paul will marry on July 6, 1996.

Original Narration

"six" or "sixth"

Time

1. Indicate time of day with numerals only. Examples: I awoke at 5:17. If you wish to attend, you must arrive by 6:25 p.m. We were expected to report no later than 1400 hours. I awoke at 4 o'clock.
2. Always use numerals when the abbreviation "a.m." or "p.m." is present. Double zeros are not necessary to indicate minutes of the hour when a whole number is used with a.m. or p.m. Examples: She leaves at 3:20 p.m. for the airport. Our hours are from 9 a.m. to 5 p.m. We're leaving at 6 in the morning.

Periods of Time

1. A decade should be captioned as "the 1980s" (not "the 1980's") and "the '50s" (not "the 50's").
2. If a decade or century is in noun form, do not use hyphens. Example:
This vase is
from the 17th century.
3. If a period of time is used as an adjective, use a hyphen. Example:
This 19th-century painting
was done by Van Gogh.

Fractions

1. Either spell out or use numerals for fractions, keeping this rule consistent throughout the media. If using numerals, insert a space between a whole number and its fraction.

Example:

Numeral Used

Do you plan
to eat 1 1/2 pizzas?

Fraction Spelled Out

Do you plan to eat
one and one-half pizzas?

2. Do not mix numerals and spelled-out words within the same sentence. Example:

Inappropriate

Malika is 13
and a half years old.

Appropriate

Malika is
13 1/2 years old.

3. If a fraction is used with "million," "billion," "trillion," etc., spell out the fraction. Example:

The population was
over one-half million.

4. Fractions expressed in figures should not be followed by endings, such as "sts," "rds," "nds," or "ths." Example:

Inappropriate

3/10ths

Appropriate

3/10

Percentages

1. Use numerals and the percent sign to indicate all percentages except at the beginning of a sentence. Examples:

Middle of Sentence

Only 6% of the votes
were counted.

Beginning of Sentence

Fifty-one percent of the people
voted "yes."

Dollar Amounts

1. Use the numeral plus "cents" or its symbol for amounts under one dollar. Examples: I need 15 cents.
2. Use the dollar sign plus the numeral for dollar amounts under one million. For whole-dollar amounts of one million and greater, spell out "million," "billion," etc. Examples: John brought only \$11. Bob brought \$6.12. The budget of \$13,000 will be sufficient. Taxes will be reduced by a total of \$13 million. He owes \$13,656,000.
3. Use the word "dollar" only once for a range up to ten. Example: I hope to find three or four dollars.
4. Use the dollar sign and numerals when captioning a range of currency over ten dollars. Example: Alice expected a raise of \$6,000 to \$7,000.

Measurements

1. Spell out units of measurement, such as "inches," "feet," "yards," "miles," "ounces," "pounds," and "tablespoons." However, if spoken in shortened form, symbols should be used. For example, if the original narration is "I'm five eight," it should be captioned as: I'm 5'8".
2. For whole numbers, use numerals. For example, caption "3 cups of sugar" instead of "three cups of sugar."

STANDARDS FOR AUDIO DESCRIPTION

The Audio Description Coalition of the United States, developed standards for audio description ([ADC, 2009](#)).

The figures [29](#), [30](#), [31](#), [32](#), [33](#), [34](#), [35](#) and [36](#) show the basic guidelines for audio description:

- Describe what you see.
- Describe objectively.
- Allow listeners to hear the dialogue.
- Trust listeners' ability to comprehend the material.
- Censorship is unfair to the material and to listeners.
- keep the language consistent.
- Race, ethnicity and nationality.
- Describe from listeners' perspective.
- Good techniques make for good description.

The figures [37](#), [38](#), [39](#) and [40](#) show the standards unique to film and video description:

- Basics.
- Race, ethnicity and nationality.
- Techniques.

Figure 29 – Audio Description Coalition - Basics guidelines: Describe what you see

Audio Description Coalition

page 1 of 31

Standards for Audio Description and Code of Professional Conduct for Describers

based on the training and experience of audio describers and trainers from across the United States

Introduction

Audio description helps to ensure that people who are blind or have low vision enjoy equal access to cultural events by providing the essential visual information. Audio description uses the natural pauses in dialogue or narration to insert descriptions of the essential visual elements: actions, appearance of characters, body language, costumes, settings, lighting, etc. Descriptions are delivered through a wireless earphone to permit people who are blind or have low vision to sit anywhere in the audience.

The Standards for Audio Description reflect audio description's origin as a means of making live theatre performances accessible; however, the spirit of these principles applies to almost all audio description situations. Other art forms and media call for variations from these original principles, which are discussed in separate sections later in this document.

The Code of Professional Conduct for Describers, near the end of this document, addresses the responsibilities of audio describers and trainers in terms of obligations to clients and consumers, privacy and confidentiality, behavior, business practices, and continuing development.

Throughout the Standards, examples appear as boxed text.

Standards for Audio Description

These standards are not intended as a do-it-yourself guide to becoming an audio describer. One should learn and practice the skills and techniques of audio description in a workshop setting led by an experienced audio description

trainer to coach and critique a beginner's first efforts. Once trained, newcomers should look for opportunities to observe the work of experienced describers and to refine their own developing abilities in peer groups of local describers. Trained audio describers are expected to deliver the best possible description.

Some suggestions for building audio description skills: Attend performances and listen to description by fellow trainees and/or veteran describers and afterwards discuss why and how they made the choices they did. Gather a small group of beginning describers and critique each other as you describe brief segments of a movie.

Movies with theatre-like pacing work best, and it's also great to use two copies of the same film—one with pre-recorded description and one without. After someone describes a segment and the group discusses it, check to see how your decisions differ from the recorded description.

Audio description grows and changes as people hone and share their skills, the material to be described evolves, and the technology for delivering description improves. In reading and practicing these standards, words like "never" and "always" must be applied with common sense. Nevertheless, the most basic principles or "rules" of audio description should not change.

Basics

Describe what you see.

- This is the first rule of description: *what you see* is what you describe. One sees physical

Figure 30 – Audio Description Coalition - Basics guidelines: Describe objectively

Audio Description Coalition

page 2 of 31

Standards for Audio Description and Code of Professional Conduct for Describers

based on the training and experience of audio describers and trainers from across the United States

appearances and actions; one *does not* see motivations or intentions. Never describe what you *think* you see.

We see "Mary clenches her fists." We do not see "Mary is angry"—or worse, "Mary is angry with John."

- Preview the material with an eye toward including the visual information that is inaccessible to people who are blind or have low vision. These include key plot elements, people, places, actions, objects, unknown sound sources, etc. not mentioned in the dialogue or made obvious by what one hears. Concentrate on that which is the most significant and least obvious from the dialogue or other audio information. Describing everything is impossible—describe what is essential in the allowable time.

Mention who answers the phone—not that the phone is ringing. It's not necessary to describe obvious sound cues.

- Describe essentials first and then, as time permits, describe further elements such as the decorative details of the settings, the physical appearance and mannerisms of the characters, architecture, clothing style, technology, color, light and texture. This sort of description works well during long pauses in the action or during scene changes.
- Description should not fill every available pause. Less is more. Audio description is not a running commentary. Listeners should be allowed to hear the emotion in actors' voices and in the tension of the silences between characters.

- Be sure to describe, as nonchalantly as possible, seemingly insignificant things the sighted audience will observe without knowing their later importance.

For example, describe that Mary is toying with a pistol and then places it in the top desk drawer. Later, when John and Mary are having a heated argument and Mary edges toward the desk, the sighted audience will suspect she's headed for the gun. By describing both actions, listeners can join in the suspenseful anticipation.

Describe objectively.

- Allow listeners to form their own opinions and draw their own conclusions. Don't editorialize, interpret, explain, analyze or "help" listeners in any other way.

This does not preclude using pre-show notes to explain something that sighted audience members easily discern but would be difficult for audio description listeners to imagine without additional information. For example, most sighted audience members see how the plant puppets in *Little Shop of Horrors* operate and then "believe" in the story's voracious plants. It's appropriate to give listeners this same opportunity.

- If the conclusion is that a character is angry, describe what led to that conclusion—the gestures/facial expressions of the character. Character's moods, motives or reasoning are not visible and, thus, not subject to description.
- Use only those adjectives and adverbs that do not offer value judgments and that are not themselves subject to interpretation.

Figure 31 – Audio Description Coalition - Basics guidelines: Allow listeners to hear the dialogue

Audio Description Coalition

page 3 of 31

Standards for Audio Description and Code of Professional Conduct for Describers

based on the training and experience of audio describers and trainers from across the United States

"Beautiful" says only that something is not ugly. But what exactly makes it beautiful? Instead of saying the person, clothing, object, etc. is beautiful, describe the things observed that caused your conclusion—so listeners may draw their own conclusion.

- It is more interesting to name the items in the clutter if time permits than to say, "The attic is cluttered." Don't take a series of specific, separate actions/events/images and describe them as one.
- When describing sizes, round-off to the next logical increment to give listeners numbers that are easier to hear and comprehend. Don't add "about" or "approximately" to qualify the estimated dimensions—this just adds words that make the information more cumbersome to grasp.

When the onstage swimming pool is 11 feet 8 inches across, 6 feet 3 inches wide, and 3 feet 7 inches deep, listeners will more easily "picture" the pool's size when they hear a less cumbersome "12 feet across, 6 feet wide and three and a half feet deep."

- Use the first person when the director has created a "first person" point of view as a means of including the audience. This sensation is part of the experience of sighted audience members, and it must be shared with listeners. This allows the describer to avoid mentioning "the audience" when, for instance, a character turns to address the audience. The same would hold true for other actions, effects, etc.

Say, "She turns to *us*" instead of "She turns to the audience." "His flashlight shines in *our*

eyes." With film and video description, the same would apply—"the shark swims toward us," not "the shark swims toward the camera" or "we move through the forest" instead of "the camera moves through the forest."

Allow listeners to hear the dialogue.

- Listeners want to hear the performance first and the description second. The dialogue is telling the story and must be heard. This rule is broken only when the confusion by omitting the description is greater than maintaining the integrity of the dialogue.

When two people are speaking at the same time, such as an actor and a describer, listeners usually cannot understand either. If you absolutely must describe over dialogue, speak loudly and insistently. Listeners, however, can usually understand simultaneous spoken words and sung lyrics.

- In most instances, a describer may talk over background music or underscoring as well as the lyrics of a repeated chorus of a song. One shouldn't describe during significant arias, the verse of a song or its first chorus. Use caution in talking over a "song played on the radio" because its recognition by the audience and/or the audience's hearing its content may be important to setting a mood, recalling an era, making an emotional statement, etc.

Deborah is talking non-stop about making a pie, but she is quietly taking a gun from a drawer. The describer must speak over her dialogue because the audience will hear a gunshot before she stops talking about making the pie.

Figure 32 – Audio Description Coalition - Basics guidelines: Trust listeners' ability to comprehend the material

Audio Description Coalition

page 4 of 31

Standards for Audio Description and Code of Professional Conduct for Describers

based on the training and experience of audio describers and trainers from across the United States

- The dialogue from radio, television or other speaking characters may be important to the story or may be considered background sound. If it is background noise, it is permissible to describe over it, assuming the description is vital.
- Don't talk in paragraphs or elaborate sentences. Use short phrases in place of full sentences. Try to speak at least two or three words so listeners have the opportunity to switch focus to the describer's voice. Unless absolutely necessary, try not to interrupt with just one word.

When there's the luxury of plenty of time: "Ken walks across the room, picks up a knife and butters the toast."

When time is tighter: "Ken picks up a knife and butters the toast."

When time is extremely tight: "Ken picks up knife butters toast."

Trust listeners' ability to comprehend the material.

- Trust listeners, who, in most instances have made the choice to attend the performance, to grasp the meaning of the material and the description. Don't condescend, patronize, or talk down to listeners.

Because the women wear bustles, the term will arise in describing the costumes. To confirm that everyone knows what a "bustle" is, tuck this into the pre-show notes without calling attention to it, with something like, "The women's long skirts puff out in back, padded over the hips and under their skirts, with bustles." If it's important to the plot, try to repeat the information during

the description for those who didn't hear the pre-show notes.

- If the play has a complex plot and/or a confusing set of characters, there's probably a synopsis in the playbill. Just as this information will be helpful to sighted audience members, sharing this information with listeners during pre-show notes may aid their appreciation of the performance and the description. Make clear that the information comes from the program so listeners understand that everyone has access to this information—that the describer is not providing special information because the listener may have trouble following the material.
- Without making assumptions that too broadly characterize the listeners, what does the describer know about their age range, their educational level, their cultural knowledge, etc.? The better one can assess listeners' common frame of reference, the better one can choose descriptions that resonate with the majority.

People who are congenitally blind (born without sight) are often comfortable with the level of information they glean from what they routinely hear and sometimes don't realize how much visual information is available.

On the other hand, people who are adventitiously blind (born with sight but lost it later) know that there's a great deal of visual information and don't want to miss it. For many diverse reasons, some people prefer minimal description for essential clarifications and others

Figure 33 – Audio Description Coalition - Basics guidelines: Censorship is unfair to the material and to listeners, and keep the language consistent

Audio Description Coalition

page 5 of 31

Standards for Audio Description and Code of Professional Conduct for Describers

based on the training and experience of audio describers and trainers from across the United States

want as much description as time allows. The describer's responsibility is to find the median.

Different listeners prefer varying amounts of description. Accordingly, the describer should honor the rules of good audio description to include the visual information that is inaccessible to people who are blind or have low vision without filling every available pause.

Censorship is unfair to the material and to listeners.

- Describers who censor information because of their own discomfort fail their listeners. Describers must say the factual information about nudity, sexual acts, violence, etc. Listeners should know everything that is evident to sighted people. If a describer feels that describing particular material will make him/her uncomfortable, s/he should not accept this assignment.

Keep the language consistent.

- Choose language that is consistent with the content of the material.
- Use language appropriate for the listeners. Children's programs should use vocabulary suitable for the age group. Make every effort to pronounce words properly—actors' names, directors' and designer's names, characters' names, the names of objects and places. Ask the theatre staff, producer, etc. for correct pronunciations when necessary. If characters' names or the names of places or things are given an unusual pronunciation in the material, be sure to use that pronunciation.

When describing for young listeners, remember that they may not have the life experience to

know common expressions like "cat and mouse game."

- Once the describer establishes a name for characters, places, objects, etc., always use that same name.
- In general, describers should avoid metaphors, similes, etc.; however, common comparisons that should be recognized by most listeners are acceptable as a means of saving time.
- Not all listeners will understand slang, colloquialisms, and regional terms. Use within context of the performance.
- Use the correct terminology so long as the majority of listeners will understand that terminology.
- Use the most descriptive words and concise phrase or sentence structure.
- Use vivid verbs. People frequently "walk" but they also amble, stagger, shuffle, saunter, and stroll. Choose the word that best matches the action.
- Use pronouns carefully. If there is only one female in a scene, then "she" is fine. If there is more than one, proper names will be clearer.
- Address time shifts (flash backs or visions of the future) in relation to the character. Music and visual effects may further identify time changes.

"Lighting shifts to pale amber as George sits next to his sister at the family dinner table."

- Use "while" and "as" to join two actions only if there is a connection between them.

"John picks up the knife as Jill turns away."

Figure 34 – Audio Description Coalition - Basics guidelines: Race, ethnicity and nationality

Audio Description Coalition

page 6 of 31

Standards for Audio Description and Code of Professional Conduct for Describers

based on the training and experience of audio describers and trainers from across the United States

- Describe colors both to help people with low vision to locate what's being described and to share the emotional "meaning" of the color in the production. People who are blind or have low vision usually share the common attributes we assign to color, such as blue and green are cool and serene while red and orange are hot and tempestuous, etc.

"The dress is burgundy" rather than "the dress is red" more richly describes the dress. Avoid, however, unusual color words: "cyan", "cerulean," "dun," "puce," etc.

Race, Ethnicity and Nationality

- Sighted audience members don't see a character's race, ethnicity or nationality; rather, they see skin color and facial features. Accordingly, the describer should simply describe each person's skin color and, if time allows, facial features. In a dramatic work where characters' backgrounds develop over time, the writer, director and actors will help the audience learn where each character fits into the world of the play—socio-economic level, educational level, relationship to other characters, race, ethnicity, nationality, etc.—to the extent that some or all of these are important to the storytelling.
- In describing *appearance*, describers should avoid offering their judgment about a character's cultural background. For example, avoid "African American" because not all people with brown skin have African heritage.
- Describing skin color and facial features is an all or nothing proposition. If it's important to note this for one character, then describe this for all characters.

- Rely on factual, clear, straightforward, commonly used terms that the majority of listeners will understand. Avoid disrespectful, derogatory or condescending terms as well as vague, poetic or euphemistic language. Do not use the terms "light-skinned" or "dark-skinned" because these phrases are sometimes associated with value judgments in the African American community.

Terms such as these may be helpful in describing skin color:

brown
dark brown
fair complexion
light brown
light tan
olive
pale pink
ruddy complexion
tan
white

- In a dramatic work where, perhaps because of the story and its setting, the characters' race, ethnicity, nationality is largely apparent to sighted audience members and integral to the plot, delineating the "sides" as part of the description would be helpful.

In *West Side Story*, the plot is much more understandable if one knows who's a Jet and who's a Shark—and that the Jets are Caucasian and the Sharks Puerto Rican. In *A Raisin in the Sun*, there would be no story without knowing that the primary focus is an African American family and that the antagonist, Carl Lindner, is Caucasian.

Figure 35 – Audio Description Coalition - Basics guidelines: Describe from listeners' perspective, and good techniques make for good description

Audio Description Coalition

page 7 of 31

Standards for Audio Description and Code of Professional Conduct for Describers

based on the training and experience of audio describers and trainers from across the United States

Describe from the listeners' perspective.

- Surprises should, ideally, come at the same time for all audience members. If characters' appearances or actions, hidden identities, costumes, sight gags, sound effects, etc. happen as a surprise to sighted audience members, don't spoil the surprise for listeners by describing (and revealing) them in advance.

If a character is in disguise, he becomes "the man" rather than "John wears a disguise." Use a neutral term "the figure in red" when characters are disguising their gender.

- If the action that accompanies a sound effect will result in a reaction from the audience, treat this as if describing a sight gag. Time the description to allow listeners to react at the same time as sighted audience members.

If the audience sees something happen that might "warn them" of the possibility of, say, a loud noise, be sure to describe that action. For instance, "Pat" loads a rifle, so we know that there's a possibility s/he will fire it.

Good techniques make for good description.

- Be certain to describe entrances and exits—who and where—especially when there's nothing audible to indicate someone has joined or left the scene.
- Use a character's name only when sighted audience members know the name. When an unknown character appears, refer to the person by a physical characteristic used in his/her initial description until his/her name is revealed. Once everyone knows the character's proper name, tie the name to the

physical description at the first opportunity ("John, the redheaded man") and afterwards use only the character's name.

Although characters are often listed in the playbill in order of appearance, sighted people probably don't remember that the first man is John, the second is Fred, and the third is Charles. So, until the character's name is used in the action, refer to the person by a physical characteristic, such as "the blond man."

- Establish and use one consistent name for each character as early as possible. This will be simpler for the describer and for the listeners.

A female character is first called "Mother," so refer to her as "Mother" until we learn another name for her. Later, other characters call her "Mary" and "Aunt Mary," but she's called the three names equally. Because, however, her greatest significance to the plot is as "Mary" (not "Mother" or "Aunt Mary"), refer to her as "Mary."

- If the characters are referred to by difficult names—long names, full names, formal names, foreign names, etc. or many variations on their names (for instance, Chekovian characters)—be sure to use the one name that best serves the storytelling.

A tip: Create a list of the established names for each character for reference during the description. A list of commonly paired couples may also be useful in plays with difficult character names.

- Once the material has identified a character, match the character's name with the actor's voice by mentioning the character's name just

Figure 36 – Audio Description Coalition - Basics guidelines: Good techniques make for good description (continued from previous guideline)

Audio Description Coalition

page 8 of 31

Standards for Audio Description and Code of Professional Conduct for Describers

based on the training and experience of audio describers and trainers from across the United States

before s/he speaks. Although the describer usually doesn't need to repeat the voice identification, this might be necessary after a character has been silent or absent for a long time or if several voices are similar and it's important to know exactly who is saying what at a particular point.

- Allow the material itself to provide information to listeners.

If a character is about to call a newly-entered character by name, or to refer to a new location by name, the describer doesn't need to provide this information. The relationships between characters may not be apparent, but it is the playwright's responsibility—not the describer's—to reveal these relationships.

- Guide people with low vision by stating the location where they should focus.

For example, in a theatre performance, say, "At the left, John enters from the kitchen." In subsequent descriptions, if the only door at the left is the kitchen door, it's fine to say, "Mary goes to the kitchen." [In film and video, say, "At the right, the hazy orange sun sets behind the mountain."]

- Match vocal delivery to the pace, energy and volume of the material. Allow the performance to set the tone and rhythm of the description, remembering that the performance, not the describer, should be the focus. Just as the describer should not assume a detached, lecturing or clinical tone, the describer should not attempt to project him- or herself into the performance as another performer. Dramatizing the delivery of the description is distracting and perhaps insulting because

listeners may feel as if the describer is telling them how to respond.

The language and delivery to describe a fight scene would differ from that used to describe a love scene.

- With experience, describers learn to gauge when laughter and applause have peaked and begun to die down. If possible, hold description until the audience begins to quiet. If not, speak loudly when describing over loud laughter, music or applause.
- When an effect will be repeated, try to describe it the first time in a way that allows a "shorthand" reference later.

In a play where characters vigorously smoke cigarettes to underscore their tension, describe the first instance as, "Mary and John light cigarettes, inhale and exhale deeply." On later occurrences, as listeners understand the pattern of their behavior, simply say, "Smoking again."

Standards Unique to Live Description

These standards apply to audio description for all types of live events, especially theatre performances. Dance and opera performances require some specific techniques, which appear following this section.

Basics

- With most plays and musicals, the describer should allow listeners to participate in the "willing suspension of disbelief" by describing in terms of the story rather than the theatrical experience. Avoid stage directions—stage right, house right, downstage, etc. As well,

Figure 37 – Audio Description Coalition - Standards unique to film and video description: Basics

Audio Description Coalition

page 14 of 31

Standards for Audio Description and Code of Professional Conduct for Describers

based on the training and experience of audio describers and trainers from across the United States

etc. may also help you keep your place if you don't read music. Be careful that looking down to read these notes does not interfere with your watching the action you need to describe.

One of the most significant elements of an opera performance is the singing—especially the famous arias. *Never* describe during an aria—unless it's to tell something of "life-or-death" importance. Just as one describes during the pauses in the dialogue of a play, the best time to describe in an opera, is during the pauses in the singing.

The other opportunity for description, if there's essential information to convey, is during a recitative. (Recitative is the narrative or dialogue parts of opera sung in the rhythm of ordinary speech with many words on the same note.)

While listeners have difficulty separating what is essentially "overlapping conversations" when presented with spoken description on top of spoken dialogue (as in a live theatre piece), spoken description over the sung words of a recitative is generally more understandable. The describer, however, should allow the singers to "establish" the chorus or recitative before interjecting the description. As always, describe judiciously, avoiding description throughout the singing.

Standards Unique to Film and Video Description

Basics

- Most audio description for film and video is scripted and recorded on an audio track separate from the material's soundtrack. This process allows the describer to write

complete, accurate descriptions that will precisely fit the available pauses. Be sure to read the script aloud at the rate it will be read for recording to verify its timing and to check for tongue-twisters. (Because the spaces for film and video description are often quite brief, the narrator usually speaks more rapidly than a describer for live performances. It's wise to rewrite to remove every possible tongue-twister.)

- The material's producer should approve the descriptions before the recording session—with the understanding that there will almost certainly be minor script adjustments to refine the timing during the recording.
- Because pauses in the material are often brief, use concise phrases or sentences. Shorter descriptions are also easier for listeners to grasp.
- If a description is essential and a silence is especially short, the describer may have to step on the first syllable or two of dialogue or narration. This often occurs when the "next voice" must be identified so listeners will understand the speaker's vantage point.
- If the musical scoring for a film or video contributes significantly to the material's emotion, try to allow listeners to be carried along with the music without interjecting descriptions. Only interrupt for vital information that must be described during the music in order to be timely.
- The narrator's voice should complement the material—it should be distinct from the voices of the characters and/or the program's narrator and mixed to sound as natural to the

Figure 38 – Audio Description Coalition - Standards unique to film and video description: Race, ethnicity and nationality, and techniques

Audio Description Coalition

page 15 of 31

Standards for Audio Description and Code of Professional Conduct for Describers

based on the training and experience of audio describers and trainers from across the United States

work as possible. The description serves the production and should blend into it.

- If there is a great deal of dialogue or narration on the soundtrack and the film or video will be shown live rather than broadcast, the producer may agree to "pre-show" description during the less important beginning credits or during an added period of pre-show "black screen."

One producer has allowed as much as 60 seconds of black screen for pre-show description. To occupy sighted audience members while the pre-show description is playing for those listening via earphone, the house lights dim to half, then fade completely out, etc.

- With a series of rapidly changing images, a technique sometimes used to establish locale or mood, briefly describe all views if possible. If not, describe the images that are most important to the plot or subject and the most visually distinguishable images to help guide people who have low vision.

Race, Ethnicity and Nationality

- Plays and full-length films or videos allow time for audiences to develop an understanding of the characters, including—when these classifications matter to the audiences' understanding of the material—their race, ethnicity or nationality. In other situations, such as brief interviews in video clips, visuals in museums and exhibits, etc., the listener's opportunity to learn about an individual is usually quite brief. If the subject matter deals with race/ethnicity/nationality or there are unexpected inclusions of people from multiple

racial/ethnic/national backgrounds, listeners will want to know this specific information about individuals' cultural heritage rather than the observable skin color and facial features.

- If sighted people will presume the race, ethnicity or nationality of a person and, thus, have a sense of "where the person's coming from," the describer should share this information with listeners. Whether the visuals or comments seem to follow racial expectations, sighted visitors are aware of the race of the person or persons depicted. Audio description listeners should have the same information.
- Check with directors, producers, etc. to determine the correct information rather than making what might be ill-founded assumptions.

In a series of homesteading photographs, the describer should identify the African American family in front of the wood frame lean-to and the Caucasian family in front of the sod house.

In a collection of video interviews about the 1957 integration of Little Rock's Central High School, sighted people are aware that this comment is made by an African American male and that comment is made by a Caucasian female.

Techniques

- Preview video the first time with no picture to help identify auditory confusions—sounds that could be mistaken for something they aren't, characters with similar voices, silent changes of scene, etc.

Figure 39 – Audio Description Coalition - Standards unique to film and video description: Techniques (continued from previous guideline)

Audio Description Coalition

page 16 of 31

Standards for Audio Description and Code of Professional Conduct for Describers

based on the training and experience of audio describers and trainers from across the United States

- Establish the location before describing anything else.

“On a downtown city street, buses, taxis, cars, bicycles and pedestrians.”

- Characters in film and video may be introduced but unnamed. Use a significant physical characteristic to identify them in descriptions.
- Without the extended time for pre-show notes, use the most revealing information about a character in place of a complete description. The age of the character may be reasonably judged by the sound of their voice. If that is not the case, mention their age. If they are dressed differently than other characters, that might be appropriate. Are they much taller or shorter than everyone else? Are they the only blonde? Do they have blue eyes, when everyone else in the family has brown?
- The relationships between characters may not be apparent, but making it so is the filmmaker’s responsibility—not the describer’s—to reveal these relationships.
- Scene changes can be confusing particularly when the soundtrack does not indicate a change. Keep them simple and short. “In the bedroom,” “at the police station,” etc.
- As well, actions, characters, and details can be confusing if we don’t know where we are. When there’s a change of place, start the description with the location, even when the resulting sentence seems awkwardly structured—the information will be well-structured for comprehension.

“In their bedroom, John and Mary embrace tightly and kiss on the lips” is better than “John and Mary embrace tightly and kiss on the lips in their bedroom”—because the preceding scene took place with the whole family gathered around the dining table and nothing on the soundtrack indicates we’ve changed locale.

- Just as a describer for a live performance should avoid theatrical jargon or references, which would draw listeners’ attention away from their involvement in the material, a film or video describer should avoid calling attention to filmmaking errors such as mistakes in continuity or editing for the same reason.
- Although one should generally avoid filmmaking jargon and describing filmmaking techniques, sometimes the brevity and simplicity of something like, “The screen fades to black” is appropriate.
- Sometimes a describer may use the word “now” to indicate a change of scene in the middle of a segment of description. Because there will be many opportunities that seem to call for the use this word, use it only when absolutely necessary.

“In the bedroom, John irons a shirt. Now, on the driveway, Mary dribbles a basketball.” In this instance, there’s a confusing change in the audio (how does ironing a shirt cause “bounce, bounce, bounce”?) Without the word “now,” do listeners assume we’re still in the bedroom and maybe hearing Mary through an open window? The simple word “now” is a great help to alert listeners that we have changed locales—just as the words that follow will help a person with low vision to focus on the new setting.

Figure 40 – Audio Description Coalition - Standards unique to film and video description: Techniques (continued from previous guideline)

Audio Description Coalition

page 17 of 31

Standards for Audio Description and Code of Professional Conduct for Describers

based on the training and experience of audio describers and trainers from across the United States

- Sometimes a describer must use a silence to describe both what's onscreen now and what's about to appear because there's no silence for the later information when it does appear. Because the audio description and the current background noise tell us we're at a waterfall, the describer needs to say "soon (or "next"), a NASCAR racetrack with a dozen cars circling the track." Without the simple word "soon" or "next" listeners would be easily confused by the describer only saying "a NASCAR racetrack ..." while the waterfall is still splashing. As well, the words "soon" or "next" alert viewers with low vision that the racetrack isn't onscreen at present.
- Occasionally there's no silent opportunity to describe something essential to listeners' understanding while that "something" is on the screen. The describer may need to omit a less significant description of what's onscreen in order to interject the critical description.

If a film about the Wright Brothers never affords a silence while one of their planes is onscreen, the describer should describe the plane when a silence presents itself and the onscreen content is less important.

- Describe the point of view when appropriate—"from above," "from space," "moving away," "flying low over the sandy beach," etc.
- We have established that we are watching film or video, so repeated references to the screen are unnecessary.
- Treat logos as any other image to be described and be sure to read the company name(s).

- Reading disclaimers and credits at the beginning and end of films, videos and television programs is an important function of audio description. As well, the describer should read text and subtitles.

Because the describer can never read as rapidly as the onscreen credits appear and disappear, work with the producer to determine the most important credits to include.

Often, some or all of the opening credits appear over the beginning of the action. In this situation, attempt to describe the action in sync with the material and read the credits before or after their actual appearance, introduce text or subtitles with a phrase such as, "Words appear" or "Subtitles appear."

Standards Unique to Museum and Exhibit Description

Audio description for museums and exhibits melds the description of visual elements with an abbreviated version of the exhibit's posted text. Visual elements include the layout/arrangement of the facility and the exhibit's components as well as the content such as dioramas, artifacts, reproductions, graphics, diagrams, drawings, illustrations, paintings, photographs, works of art, maps, video kiosks, computer interactive stations, etc. If the exhibit includes touchable objects, the description would also guide users' exploration of these items. Text elements include the titles, explanatory text, quotations, legends, and captions.

Basics

- In order to provide an experience similar to that of sighted visitors, audio description users