Web prototype for creating descriptions and playing videos with audio description using a speech synthesizer

Sebastian Calvo-Salamanca
Systems Engineering Program
Bogotá, Diagonal 46 A # 15 B – 10, sede Claustro
+57 1 3277300
scalvo22@ucatolica.edu.co

Andrés Felipe Coca-Castro
Systems Engineering Program
Bogotá, Diagonal 46 A # 15 B – 10, sede Claustro
+57 1 3277300
afcoca03@ucatolica.edu.co

John Alexander Velandia-Vega
Systems Engineering Program
Bogotá, Diagonal 46 A # 15 B – 10, sede Claustro
+57 1 3277300
javelandia@ucatolica.edu.co

ABSTRACT
The audio description mechanism allows people with visual disabilities to access video content. Currently, commercial and open source tools exist with features such as creation of descriptions and play videos with audio description which are based on human voice recording. An alternative Web prototype is presented in this research with the aim of introducing a novelty way for accessing video content using a speech synthesizer. Assessment of the proposed prototype is accomplished from user experience perspective and accessibility perspective. It is considered a sample of participants with visual disabilities to evaluate the user experience. Outcomes from the assessment stand out that the prototype is intuitive and easy to use according to the Average System Usability Scale. Thus, functional users have great user experience when they use the prototype for playing videos with audio description. As general conclusion, it is concluded that the proposed prototype is a potential alternative tool for creating descriptions and playing videos with audio description.

General Terms
Measurement, Documentation, Performance, Design, Reliability, Experimentation, Human Factors, Standardization, Languages, Theory, Audio description.

Keywords
Accessibility, audio description, text-to-speech, speech synthesis, social inclusion, Web content, blind, visually impaired.

1. INTRODUCTION
Improving quality of life of people with visual disabilities is a technological and legal issue [6]. Namely, guaranty the accessibility to information systems has become an important quality attribute that public information systems must contain implicitly [7].

People with visual disabilities tackle with unbreakable barriers when they access Web platforms that provide audio visual information, leading them to social exclusion, since they are being excluded from accessing the Web content they want to interact with [16].

2016 8th Euro American Conference on Telematics and Information Systems (EATIS), Cartagena de Indias – Colombia. April 27 - 29.

Video audio description is a technique that attempts to compensate the lack of visual perception, so that a person with visual impairment or any visual disability perceives the video message as if he were a person with none visual impairment or visual disability [13]. There are two sort of video audio descriptions: (1) standard which is played while the video is running, and (2) extended which consists of having a paused video, as long as the audio description is played, once the audio description finishes the video starts off from the last point where it stopped.

The objective of this paper is to answer the following question: Which alternative tool for creation of descriptions and playback of audio described videos could be provided to people visually impaired or with any visual disability to guarantee the access to video content? Videscribe is the proposed prototype that is designed, developed and tested to address this research question.

The main contribution of this paper is that throughout an engine based on a speech synthesizer, audio descriptions for videos are created to guarantee accessibility to people visually impaired or with visual disabilities, with novelty of eliminating human voice.

The remainder of this paper is organized as follows. An introduction to audio description topic and its components is presented in Section 2. The methodology addressed in this research is contained in Section 3. Section 4 sets out the design of Videscribe, involve processes, deployment, and some others diagrams that help to describe the architecture of the prototype, which are also tied to the implementation phase. The evaluation of Videscribe is presented in section considering user experience perspective and accessibility perspective. Outcomes from the assessment are presented using mathematical approach, section 6. Analysis of the outcomes and the conclusions of this research are found it in section 7. Section 8 provides to future researches challenging topics from which new studies and explorations may come up.

2. AUDIO DESCRIPTION AND TEXT TO SPEECH
Audio description is defined as an inter-semiotic translation from which visual scenes are transferred into words, then these are received aurally by end users[13].

It is a mechanism oriented to users who are blind or visually impaired that allows access to visual information appearing on screen, which they would otherwise miss[11], that audio descriptions precisely timed to occur only during the pauses in dialog or significant sound elements or performing arts or in media allows persons with visual impairments to have a greater access to the images integral to a given work art. Audio description has been shown to be useful for anyone who wants to truly notice and appreciate a more full perspective on any visual event.

The audio description with TTS is an audio description from which instead a human narrator, the description is read by a Text-To-Speech software, and it has advantages like lower production cost, it does not need previous recording of the descriptions to load the descriptions and is not required a high speed internet connection because the descriptions are in text files[4].

Text-to-speech audio description has several advantages. From the perspective of the audio description provider, TTS AD offers unequalled cost-effectiveness in terms of AD production in comparison with conventional methods of producing audio description. TTS AD does not require the recording of the AD script (for pre-recorded AD), nor does it incur any human labor costs for the reading out of the AD script (for live AD). Furthermore, in contrast to audio describers involved in the production of conventional AD, who need to be able to develop "the vocal instrument through work with speech and oral interpretation fundamentals"[4].

3. METHODOLOGY

A Web prototype was developed according to a Participatory Action Research (PAR) methodology and Scrum framework. PAR is a research method involving both participants and researchers throughout the process from the initial stages to gathering and communicating final results[3]. In this research, PAR was performed over visually impaired population. Community contact was initially established with the support of a social informatics expert who provided the contact of IT promotion head from the National Institute for Blind People (INCI, Spanish). This person was asked about the barriers that target population tackle when they use technology. The main problem identified was the lack of tools for creating descriptions and playing videos with audio description. The Web prototype was proposed and developed as a solution to this problem considering the IT promotion head expertise. Finally, an assessment to the Web prototype was conducted with visually impaired people to measure its usability and performance playing a video with text to speech audio description. This assessment was implemented through a survey contained multiple-choice Likert-scale questions [5].

Scrum, an incremental and iterative hybrid framework, was used to develop the Web Prototype [15]. This development considered the phases of requirements definitions, design, development and the Web Content Accessibility Guidelines (WCAG 2.0) for prototype compliance. First, functional and non-functional requirements definitions were set in a product backlog of eighteen user stories allocated in four categories: authentication, audio descriptions management, video audio description and quality attributes. Then, in the design phase, main processes supported by the Web prototype were designed using the Business Process Management Notation. Web prototype mockups, component diagram, deployment diagram and logic architecture were done within the same phase. Next, the Web prototype development was performed using the server-side PHP framework Codeigniter that operates with Model-View-Controller pattern design, REST services and client-side framework AngularJS. Text to speech audio description was achieved 1) driving the text to the speech synthesizer; 2) returning the audio description to the client-side; 3) and synchronizing it with the screened video. Finally, an accessibility assessment of the Web prototype was performed to verify its fulfillment according to both levels A and AA of the WCAG 2.0.

4. WEB PROTOTYPE

Videscribe is a Web prototype that creates audio descriptions and plays videos using audio descriptions throughout text-to-speech technology. The design and implementation sections encompass the processes, the software architecture and the models that Videscribe supports.

4.1 Design

4.1.1 Business Process Model

Business Process Management Notation (BPMN) is used to model the activities and the processes that Videscribe supports, since this notation offers a formal modeling of workflows processes[2]. Figure 1 and Figure 2 depicts throughout BPMN the principal features of Videscribe, create and play audio descriptions.

Figure 1 describes the process of creating and editing audio descriptions. The involved activities consist of writing the title of the audio description, which serves as identifier for every video. Afterwards the audio description language is set out according to the needs of users. The followed activity lie in selecting a URL where the video is hosted, for instance YouTube, then a validation is executed to ensure the URL's correctness. In case the URL is not valid, the user has to insert again a valid URL, otherwise the video is loaded into the prototype. Then, the prototype adds fragments to be incorporate into the audio descriptions, which become an input for the text-to-speech. Finally, users are able to play or delete the audio description, which is confirmed sending the Web form.

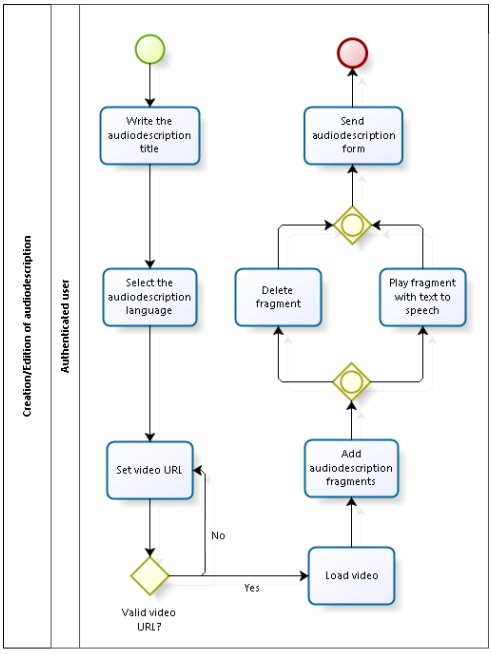

Figure 1. Process to create and edit an audio description

Figure 2 describes the process of playing videos by the user. The first activity consists of synchronizing both the prototype and the video, thus once the video is running the fragments are also being performing. Then, the user is able to accomplish any of the following activities: obtaining the URL of the audio description and insert code, this option allows the user to insert the video including the audio descriptions into another Web site, avoiding the user to visit again the Videscribe platform. The user also has the choice of embedding the video into another Web site more transparently.

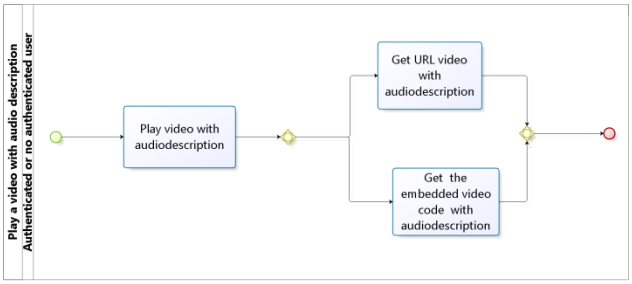

Figure 2. Process to play a video with audio description.

4.1.2 Deployment model

This sort of model intents to describe the physical and logical components that are involved in the interaction between the prototype and users, including their relationships [9]. Figure 3 depicts a server and client architecture, from which Hyper Text Transfer Protocol (HTTP) is used to communicate them. Two tiers conform the physical deployment model, first tier encompassed the server, and the second tier is part of the client.

The server is composed of Windows 2008 as operating system which at the same time hosts Tomcat as Web Server and the database management system MySQL. Videscribe uses the default Tomcat´s libraries. Videscribe could be deployed in versions 2.1 or higher. The client could be deployed on any operating system that runs any Web browser, for instance Google Chrome.

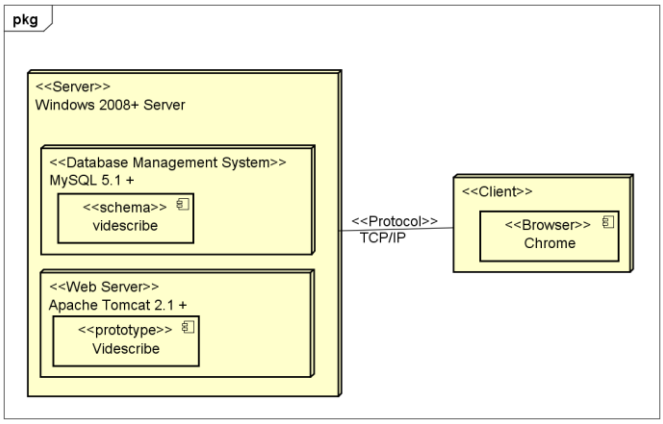

Figure 3. Deployment diagram

4.1.3 Logical model

The software architectural pattern used for implementing Videscribe is the Model View and Controller (MVC), because it has benefits such as separations of concerns in the code base avoiding mixing code, developer specialization, and components´ reusability [8].

The flow of actions of this architectural pattern starts when a user makes a request over the Web browser to the controller. The controller takes this request and the model makes operations according to execution parameters that are send by the controller [10].

The controller gets all the resulting information from the operations and loads the correspondent view with this information in a visual representation, commonly with a HTML document. Finally the controller responds with this visual representation to the browser to be presented to the user (see Figure 4).

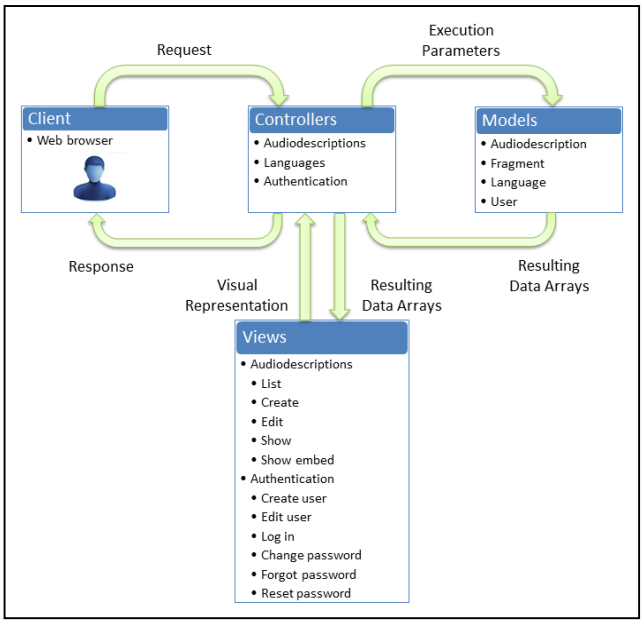

Figure 4. Flow of actions Model-View-Controller

4.2 Implementation

4.2.1 Proposed software architecture

The implementation is based on two logical layers, the client and the server. On the client side lies in the presentation and business logic layers, it means that most of the processing is performed on the client side, having rich client features, independently of the server. The server side incorporates a Web service interface and model layer to access the physical database.

Client side comprises three components: *(1) views* that is a set of Web pages. *(2)AngularJS* which is a JavaScript framework for dynamic web apps, it is used for building up Rich Internet Applications (RIA). *(3)* The last component is the text to speech *engine*, which is the novelty of this research, it embraces the following components:

- *Video descriptions*: It is in charge of defining the audio descriptions for every video.
- *Call TTS*: It is an interface that is triggered by the user interface and it has the function of calling the text to speech component.
- *AD:* It is the audio description component which is responsible for executing the audio descriptions associated to the video.
- *Video player control:* This component has the aim of playing, stopping and in general de handing of videos.

- *Sync AD with video:* The objective of this component is to synchronize the video with audio description while video is executing. The process to synchronize consists of executing the video player JW Player which subsequently search for the storage where the video is hosted, normally YouTube, afterwards the video player loads and plays the video, and at the same time the responsive voice is synchronize with the video.

Server side provides a restful Web service interface that incorporates the CRUD (create, retrieve, update and delete) operations, guarantying loose coupling among components. Once the request is received by the server, Videscribe interacts with the database to handle data and to provide the response to the client, namely with AngularJS.

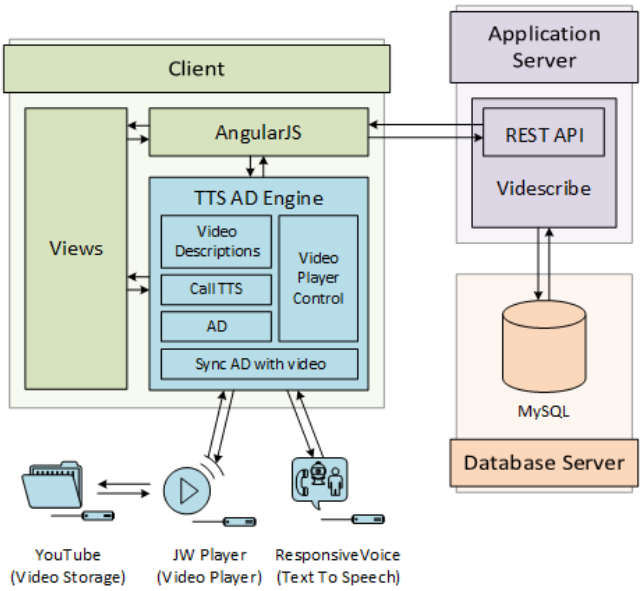

Figure 5. Proposed software architecture

4.2.2 Essential graphical interfaces

This section presents the main Web user interfaces that have been implemented for Videscribe, these interfaces comprises features such as the creation, editing and some other operations that users may perform over the prototype.

4.2.2.1 Edition of audio description

As requirement, the user has to be logged into the information system to perform any task over this Web page. In the edition page users has the choice of changing the title, the audio description language and the video's URL. Once the video is loaded, fragments for describing video scenes might be added by the user. It is possible to edit, play or delete the fragments, change the initial time of the fragment, and choose if the fragment is standard or extended audio description, Figure 6 depicts the features previously mentioned.

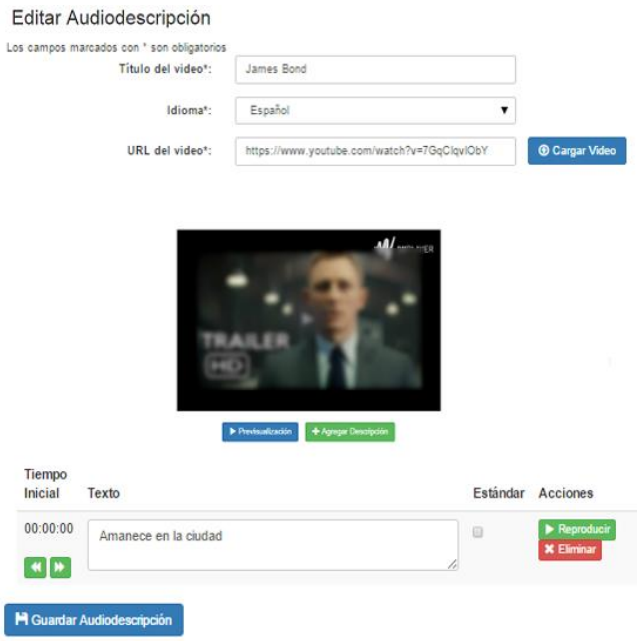

Figure 6. Edit audio description page

4.2.2.2 Play video with or without audio description

In the Web page, namely on the playback component the user might play the video with or without audio description. It is possible to pause and play the video at any point of time. Moreover, it is possible to skip forward or skip backward the video playback, Figure 7.

The concept of play the video with or without audio description looks for the social inclusion term to be accomplished because the people with or without visual impairment can make use of the prototype.

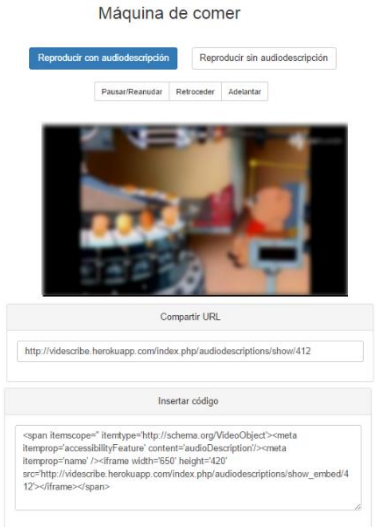

Figure 7. Playback page

4.2.2.3 Main page

The user may or may not be logged-in to see this page. This page exposes all the audio described videos (see Figure 8).

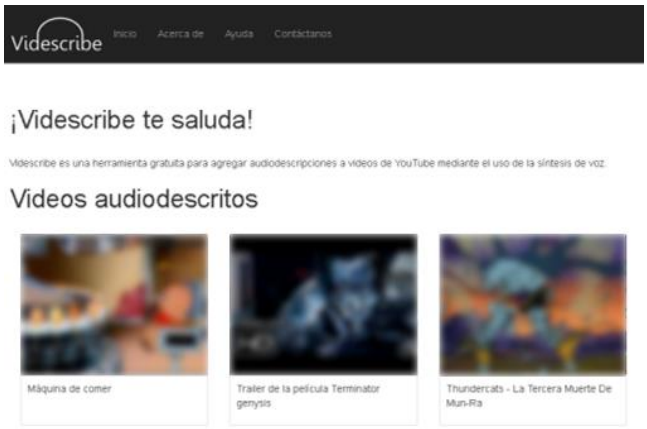

Figure 8. Main page

4.2.2.4 List of audio descriptions page

The user has to be logged-in to use this page. In this Web page logged users are able to manage their audio described videos. Additionally, users might navigate to another Web pages to create, play, edit or delete audio descriptions, see Figure 9.

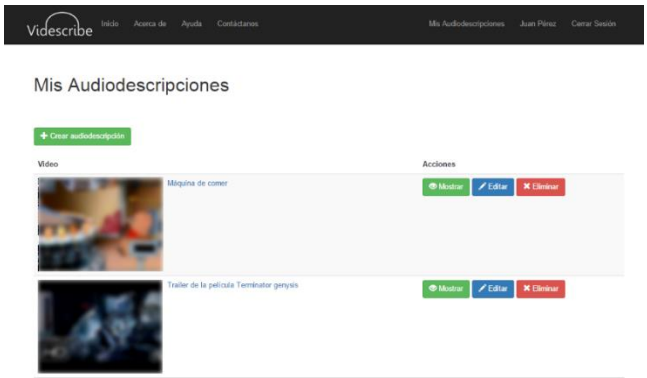

Figure 9. List of audio descriptions page

5. PROTOTYPE ASSESMENT

The proposed prototype is assessed from two perfectives: user experience, from which a sample of people is taken to test features provided by Videscribe, Likert scale is applied to design and develop the questionnaire that allows assess this perspective. The goal of the second perspective is to address the accessibility quality attribute, which is evaluated using Web Content Accessibility Guidelines (WCAG) as set of recommendation of the World Wide Web Consortium (W3C).

5.1 User experience

The participants were a sample of the visual impairment population that wants to access to the Web content. The data collection instrument was applied to a total of 36 participants, 12 of female gender and 24 of male gender. The average age of the participants was between 18 and 29 years old.

The target population of the prototype was formed by the people with visual impairment that want to access to videos using the Web. The sample corresponds to a no probabilistic sample of volunteer participants.

A questionnaire is presented to measure the usability and experience of people with visual impairment at the moment of playing a video with audio description.

The questionnaire is composed of participant identification items and items with Likert scale where 1 is totally disagree and 5 is totally agree[5], to know the experience in the reproduction of a video with audio description and measure the prototype usability, and it has three parts: the participant recognition, the playback experience with audio description and finally the prototype usability.

The procedure was first collect the participant information, such as gender, education, kind of visual impairment and experience with audio description and speech synthesis.

To measure the experience on the playback of an audio described video, a fragment of the movie *The Pursuit of Happyness* was chosen and was audio described using the prototype. To do this measure the participant played the audio described video that is located on the Web page of the Web prototype, the participant had to click on the button play with audio description, and at the end of the playback is asked to the participant to answer the items related to the last part.

After this step, is asked to the user to answer the items that lead measure the prototype usability.

At the end of the questionnaire is presented a declaration of consent, in which the participant agrees to participate in the investigation and a section to leave comments.

5.2 Prototype accessibility

The prototype was implemented following the Web Content Accessibility Guidelines (WCAG) that is a stable, referenceable technical standard. It has 12 guidelines that are organized under 4 principles: perceivable, operable, understandable, and robust. For each guideline, there are testable success criteria, which are at three levels: A, AA, and AAA [14].

An accessibility evaluation was realized to the main page and video playback page according the WCAG 2.0 standards. That evaluation was realized to verify the accomplishment of the conformity requirements of the levels A and AA.

The main page meets the 64% of the conformance requirements of the A level and 34% of the requirements are not applicable. The main page meets with the 100% of the conformance requirements.

The playback page meets the 72% of the conformance requirements of the A level, the 4% no meet the requirements and the 24% is not applicable. The playback page meets with the 94.74% of the conformity levels of A level.

6. RESULTS

6.1 Questionnaire

The questionnaire was organized in the following way:

- Participant recognition.
- User experience with the reproduction of a video with audio description.
- Prototype usability

6.1.1 *Participant recognition*

In this part of the questionnaire a set of information was required to identify and recognize the participant.

The average kind of visual impairment of the sample shows that the most of the sample is blind people with the 89% (see Figure 10).

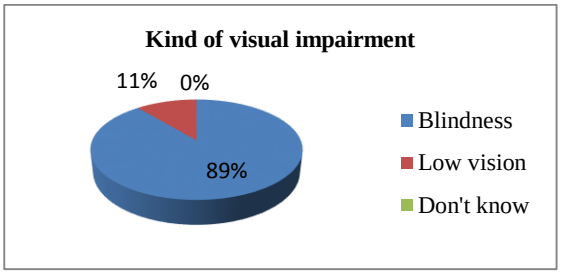

Figure 10. Average kind of visual impairment

6.1.2 *User experience with the reproduction of a video with audio description*

In this part of the questionnaire is asked to the participants about the experience at the moment of playing a video in the Web prototype. A high percentage of the participants determined that the content presented in the video was accesible with a total of 69%. The minor cipher were the people that were totally disagree with the affirmation the content presented in the video was accesible with only 3%.(see Figure 11).

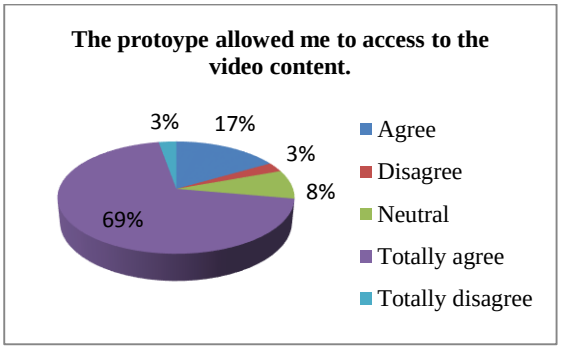

Figure 11. Video content accessibility.

6.1.3 *Prototype usability*

In this section the System Usability Scale (SUS) method was used to know the usability scale of the Web prototype, the SUS is a simple, ten-item scale giving a global view of subjective assessments of usability [1].

The first affirmation of this section was “I think I would like to use this prototype frequently”, where the highest percentage was 28% that answered that they would use the prototype frequently, on the other hand the participants that were totally disagree with that affirmation was a total of 6% (see Figure 12).

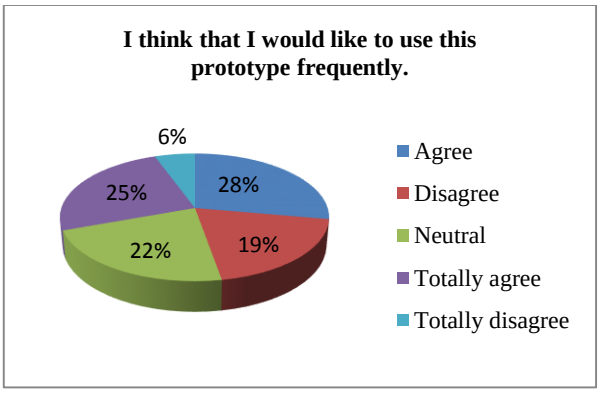

Figure 12. Frequent use of the prototype

61% of the sample thought that the Web prototype was easy to use. On the other hand with 3% are the participants that though that the prototype was difficult to use and the participants that disagree with the easy way to use the prototype (see Figure 13).

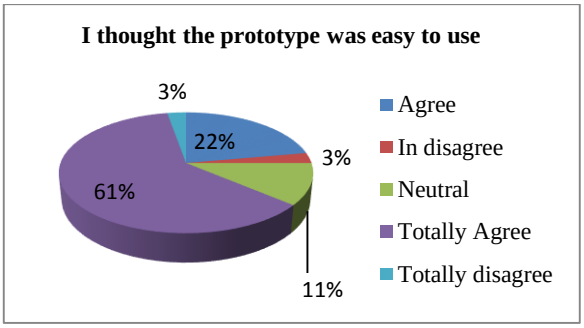

Figure 13. Facility to use the Web prototype.

7. DISCUSSION AND CONCLUSIONS

In the results sections in this paper is possible to see that 61% of the sample though that was easy to use the prototype, that means that one of the goals of the project its being fulfilled.

The Web prototype is a potential alternative tool for creating descriptions and playing videos with audio description because it allows access to the content of videos according to 86% of the participants.

Web prototype provided a good experience score of 23.47 in a 7 – 35 scale to subset of people with disabilities in playing a video with audio description due to results obtained

Web prototype has a greater usability than average SUS score due to the score obtained by applying SUS.

Implementation of the Web prototype is based on user stories and design defined. The evaluation results show that the Web prototype was developed as it was specified.

8. FUTURE WORK

The future work should include add the YouTube search to the prototype, avoiding the need to copy and paste the YouTube URL. Including additional features to the prototype would make it more accessible to people with other disabilities for example subtitles for deaf people, having account the people with or without disabilities, it is necessary to increase the accessibility and usability level to make the user experience more comfortable.

It is also proposed to add a search tool to find the audio described videos, adding the categories section to make the search easier to users. Implementing a virtual tour guide to create, edit and delete audio descriptions in the Web prototype would be also a help for blind people.

9. REFERENCES

[1] Brooke, J. 1996. SUS - A quick and dirty usability scale. *Usability evaluation in industry*. 189, 194 (1996), 4–7.

[2] Chinosi, M. and Trombetta, A. 2012. BPMN: An introduction to the standard. *Computer Standards and Interfaces*. 34, 1 (2012), 124–134.

[3] Comeau, S. 2014. *Participatory Action Research: An educational tool for citizen-users of community mental health services*.

[4] Fernandez-Torne, A. and Matamala, A. 2015. Text-to-speech vs. human voiced audio descriptions: a reception study in films dubbed into Catalan. *Journal of Specialised Translation*. 24 (2015), 61–88.

[5] Hartley, J. 2014. Some thoughts on Likert-type scales. *International Journal of Clinical and Health Psychology*. 14, 1 (2014), 83–86.

[6] Lazar, J. et al. 2004. Improving web accessibility: A study of webmaster perceptions. *Computers in Human Behavior*. 20, 2 (2004), 269–288.

[7] Lazar, J. et al. 2012. Investigating the Accessibility and Usability of Job Application Web Sites for Blind Users. *Journal of Usability Studies*. 7, 2 (2012), 68–87.

[8] Pop, D.P. and Altar, A. 2014. Designing an MVC model for rapid web application development. *Procedia Engineering*. 69, (2014), 1172–1179.

[9] Process, R.U. 2000. *Visual Modeling with Rational Rose 2000 and UML TERRY QUATRANI Publisher : Addison Wesley Second Edition October 19 , 1999 ISBN : 0-201-69961-3 , 288 pages*.

[10] Shalloway, A. and Trott, J. 2002. Design Patterns – Elements of Reusable Object-Oriented Software. *A New Perspective on Object-Oriented Design*. (2002), 334.

[11] Snyder, J. 2007. Audio Description: The Visual Made Verbal. 2, (2007).

[12] Strauss, A. and Corbin, J. 2008. *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*.

[13] Szarkowska, A. 2011. Text-to-speech audio description: towards wider availability of AD. *The Journal of Specialised Translation*. 15 (2011), 142–162.

[14] Termens, M. et al. 2009. Web Content Accessibility Guidelines : from 1 . 0 to 2 . 0. December (2009), 1171–1172.

[15] Yang, C. et al. 2016. A systematic mapping study on the combination of software architecture and agile development. *Journal of Systems and Software*. 111, (2016), 157–184.

[16] Zaphiris, P. and Ellis, R.D. 2001. Website Usability and Content Accessibility of the top USA Universities. *WebNet 2001 Conference*. 2001 (2001), 1–6.