In-Situ Study of Blind Individuals Listening to Audio-Visual Contents

Claude Chapdelaine
Vision and Imaging, CRIM
405 Ogilvy Avenue, Suite 101
Montreal (Quebec) Canada
(514) 840-1234
Claude.Chapdelaine@crim.ca

ABSTRACT
Videodescription (VD) or audio description is added to the sound track of audio-visual contents to make media such as film and television accessible to individuals with visual impairment. VD translates the relevant visual information into auditory information. In our previous users' testing, we found that the need of VD could be quite different depending on the visual disabilities of the participants. In order to better identify those differences, we conducted a study with ten legally blind individuals (with and without residual vision) to observe the type, quantity and frequency of the information needed by them. We learned that the degree of residual vision and the complexity of the content have a significant impact of the required level of VD. This suggests that a tool to render VD should offer a basic level of information, allow enough flexibility to provide more VD if needed, and answer on the fly demands for specific information. These specifications were implanted into an accessible video player.

Categories and Subject Descriptors
H.5.2 [**User Interfaces**]: Evaluation/methodology; K.4.2 [**Social Issues**]: Assistive technologies for persons with disabilities.

General Terms
Human factors.

Keywords
Multimedia, accessibility, audio description, blind and visual impairment.

1. INTRODUCTION
Accessibility of audio-visual content such as television and film, to the visually impaired individuals (VII), is rendered through added audio information describing the relevant visual information. This added information is called audio description or videodescription (VD). Since 2004, CRIM has been developing tools to produce or render VD. Our work in production aims at minimizing production time, while our work on rendering is done through the implementation of an accessible VD player. The goal of all our developments aims at integrating a comprehensive understanding of the users' cognitive and memory capacities.

ASSETS'10, October 25–27, 2010, Orlando, Florida, USA.

When listening to television and film either for learning or entertainment purposes, VII are required to process audio-visual information according to their available resources such as having residual vision or not. In this context, they must strongly rely on the audio channel and without the proper translation of the visual element, information is missing. The effort made by them to "fill in the blanks" creates a cognitive overload that often results in task abandonment. The objective of this study is to assess the nature of the information missed by the intended users in order for them to understand the visual component of the content. Our approach is to conduct a four phases' study to observe the type, quantity and frequency of the information needed by the participants who are legally blind (phases I and II) and others who have low vision (phases III and IV).

We have completed phases I and II of this study that involved ten legally or totally blind individuals (with and without residual vision). They are either congenitally or late blind individuals who are considered in levels 3, 4 and 5 of the World Health Organization (WHO) [1] classification. Phase I included results observed while participants listened to a short film (romance) and a TV drama [2]. For phase II, participants viewed a more complex film (action movie) and a scientific TV magazine. Phases III and IV will present the same data to participants with low vision, corresponding to levels 1 and 2 of WHO. Our ultimate goal is to develop tools that would offer accessible and adaptable VD so that VII could autonomously experience an understandable and enjoyable audio-visual content.

In this paper, we report on the combined results for phases I and II. Section 2 presents the specific cognitive capacities of VII. In section 3, we describe the methodology used in the study, in terms of procedure used, profiles of the participants and the dataset presented. Section 4 elaborates on the results obtained and section 5 discusses the impact of these results on the design of the player.

2. COGNITIVE CAPACITY OF VII
The WHO defines five levels to classify the spectrum visual disabilities which are based, as shown in table 1, on visual acuity and/or visual field of view (FOV) in the better eye with the best possible correction [3].

For people in levels 4 and 5, access to information requires a mandatory auditory or tactile channel. Light perception and the viewing of high contrasting forms characterize level 4 while level 5 corresponds to no vision. At these levels, computer usage requires mandatory help of a screen reader or a Braille display [4]. As for TV, content will be best understood if auditory input is clear and fluid without conflicting audio messages. At level 3, people have a modest functional residual vision. The disability is

qualified as severe low vision and individuals could be classified as blind or visually impaired.

Individuals with functional residual vision are in level 1 or 2. They are usually referred to as persons living with low vision. Illumination and contrast are the most vital factors to maximize the use of their remaining vision [5]. For levels 1, 2 and even more for level 3, the visual channel must combine auditory and tactile inputs to successfully complete a task. For example, using a computer will require a large screen with magnification, and a voice synthesizer could be needed depending on different factors such as document length, fatigue level, etc. [4]. As for TV, they will need to be close to the screen with reduced illumination and good contrast. For them, the auditory channel is important and the content will be easily followed if visual input is without conflicting messages.

Table 1. Five levels of WHO's classification

Level	Degree of acuity	FOV
1	> 20/70	> 60 degrees
2	> 20/200	> 20 degrees
3	> 20/400	> 10 degrees
4	> 20/1200	> 5 degrees
5	None	None

Defining FOV, an acuity classification for VII is useful for statistical purposes and crucial to rehabilitation and service access. However, the frontiers between levels are not always clear since the same impairment can cause different inabilities while different impairments can induce the same inability [3]. Also, other factors have an incidence on the person's capacities such as the age at which the disability started if not congenital. In a functional MRI study, Sadato et als, [6] found that individuals who became blind before the age of sixteen could redirect their primary visual cortex association from visual input to tactile input.

The degree of remaining vision and the rehabilitation of the VII will have a tremendous impact on his/her ability to process information. For everyone, the human sensory system has limitations when processing quantity and quality of the information which has a direct impact on attention resources. Colavita [7] established the visual dominance effect, where humans show a strong tendency to rely more on visual information in a multichannel environment. However, when this predominance can not entirely be used, as for VII, humans must rely on other channels such as the auditory channel. But this channel is different from the visual one and cannot be used in the same way. Audio is omnidirectional and transient [8]. Omnidirectional implies that sound can be heard coming from any direction as opposed to the viewing of the object that necessitates the direction of the eyes. This makes sound more prone to distraction than viewing and would require more attention to keep the focus. Sound is also limited in time, thus transient, as opposed to text or image that can be looked for any amount time. Understanding audio requires more preattentive processing; making it more cognitively demanding than vision. This necessitates more use of short term memory. The very nature of the audio modality has a huge impact on the cognitive demand of the VII.

Research in cognition shows how visual impairment modifies the way human processes information and that auditory, tactile and kinesthetic channel will become central [9]. Brain plasticity research demonstrates how visual deficiency changes the demand to the brain [10]. It proves that even when the loss of vision is temporary, the brain quickly adapts to that change. Prolong vision loss and especially early blindness can generate new neural connections [11]. Vanlierde [12] suggests that cross modal brain plasticity enables the auditory and tactile channels partly used to generate a vision in the brain. Rokem and Ahissar [13] found that congenitally blind individuals have higher auditory and memory capacities than sighted people. They are more resilient to noise and better at frequency discrimination which allows them to augment speech perception. They suggest that this better encoding of audio could explain the short-term memory capacity advantage of the blind individuals. Surely, these cognitive limitations and advantages will play a crucial in scripting an efficient VD.

In our prior work [14] [15], we tested film with VD produced according to the existing guidelines. The feedback of VII to whom films with VD were presented, suggested that different quantity of VD would accommodate a broader range of vision disabilities and individual preferences. For instance, in testing sessions where we presented film with VD available commercially, the VD was often rich with an elaborate choice of words. From this viewing, frequent comments were that 1) there were no moment left to hear the ambiance of the film and 2) it was tiresome to be so attentive to everything that was being said. In Chapdelaine [16], who presented to VII, films with two levels of VD (standard and extended), the ones with low vision and the congenitally blind individuals stated that they needed less VD and preferred the standard level. On the contrary, late blind individuals preferred the extended version. Furthermore, individuals with more residual vision reported that they found annoying or confusing when the VD was not synchronized with the occurrence of the event. Those results suggested the need for a user-oriented study that could identify the criteria of an efficient VD in a real life context and how those criteria could be applied to satisfy the various needs of the intended population. But first, we need to understand how VD is actually produced and know what the findings from research on VD are.

3. THE NATURE OF VD

To produce VD, scripters composed text descriptions to translate relevant visual elements so that VII can follow the storyline. The scripted VD is rendered in audio and added to the content. The VD is inserted in the gaps between dialogues and is synchronized as closely as possible to the related event. The limited number of those gaps and their short duration imposes many complex constraints on the production. Adding VD to an existing audio track that is rich in dialogues, ambient music and relevant audio sounds constitute a challenge. Thus, this complex task implies many editing choices for which only few guidelines are available [17][18][19]. Those guidelines are often based on intuition or convention without any indications on why some VD may be more effective than others [20]. But in general, the scripters tend to present as much VD as they can which, according to user's feedback, may not be what they need. Thus, VD producers would greatly benefit from more comprehensive guidelines based on user-oriented research.

Recent research on VD could be divided in two fields: 1) the informational value for indexing and classification and 2) the

linguistic nature of VD. Turner proposed a VD typology to augment film indexing [21] and to automate VD production [22]. From 11 different audio-visual contents, he classifies information to establish the most frequent categories observed in VD. The categories identified formed a typology of VD composed of action/movement, character identification, description of the surroundings, expressions of emotion, and textual information included in the image. On the linguistic approach, research aims at understanding its linguistic nature and how visual cues could be translated into words for VII. Piety [23] demonstrated how the constraints imposed on VD production create a distinctive usage of language that has its own form and function. He studied not only which visual cues the producers choose to translate into VD, but also how it was described. He found that particular form of language used had an impact on the cognitive load of VII. Salway [24] showed a relation between the frequency of words used in VD and the occurrence of the characters, the action and the scene. He found that the observable degree of regularity of words in the VD corpus might facilitate the automatic production of VD. Peli [25], Pettit [26], Schmeidler [27] and Ely [28] reported on evaluations done by the VII on the value and importance of VD. Despite the provision of these works to better understand the nature and role of VD, still little is known on how to formulate an effective VD that would transmit the necessary information and assure a good comprehension of the visual message [23].

4. STUDY METHODOLOGY

This methodology was inspired by the theory of Suchman [29] on situated cognition which points out the role of the environment in the cognitive process which she studied using methods such as verbal analysis. Since, VII already watch film and television where their relatives would supply the necessary information. We opted to study those conversations which took place in real-life conditions (as much as possible) and without any prior training of the participants.

4.1 Procedure

We conducted the phases I and II of an in-situ user study based on verbal analysis. The design scenario created a realistic context of VII watching television. The participants would watch in a room of their choice, audio-visual contents with a sighted person (experimenter). Participants were told to seek information from the experimenter as they usually do with a relative. A brief synopsis of the content was read to the participants before viewing. Participants could ask questions before and during viewing whenever information was needed. After the viewing, participants were asked to summarize what they viewed as if telling the story to a friend. The recollection was used to build the mental representation of their comprehension. If concepts were omitted, the experimenter would ask a related question to know if the concept was understood or simply omitted.

The sessions lasted on average an hour and a half. The required viewing time for the dataset without questions was 30 minutes and the summarization took on average 15 minutes. This left an average of 30 minutes for the information request during viewing time. The sessions were filmed and the recording was used to analyze the verbal protocol from which we extracted all the requests made by the participants and a transcript of the summarization. The requests were classified into two groups. Group 1 included requests that were made to confirm information. For example: “Is this Marc speaking?” or “Is the man sitting next to the woman?”. Group2 included the requests that were an inquiry to get information, such as: “Where is this happening?”, “What is the man doing?”. Both groups were further classified into of six categories of request: who, where, action, facial expression, description, identification of sound or speech.

The summary made by participants and their answers to the questions of the experimenter were used to build a mental representation. The landscape model approach was used based on the recommendation of Roskos-Ewoldsen et al. [29]. They confirmed the adequacy of the model to describe the mental representation of TV series. The extracted concepts from the model of each participant (summarization) were compared to the concepts collected from the models of four sighted persons (control group) who summed up what they remembered from their viewing. The inter-reliability among sighted viewers on concept identification was 94% after discrepancies were resolved.

4.2 Participants

The individual sessions were done with ten participants in a room of their choice (home, rehabilitation center or work). Results were later divided into two groups. Group A included five individuals who could be classified at levels 4 and 5 (2 congenitally blinds, 2 late blinds before the age of 16 and 1 late blind for more than 10 years). Individuals in group A were between 36 and 65 years old and they all reported listening to less than five hours of television per week. All of them except one stated that they rarely watch television alone. Three of them preferred less VD while the remaining two were satisfied with the actual level.

Group B was composed of five legally blind individuals with some residual vision (level 3). They reported being able to detect a human face, some of them could identify a movement done by one person but not if it was a group. They all needed to be very close to the screen and stated that image contrast played a very important role in their ability to identify anything. Most of them dimmed the light in the room. Their age varied between 46 and 65 years old. Three of them often watch television alone while the two others rarely do so alone.

Participants of both groups had experiences with VD and they all stated preferring VD to a human reporter for autonomy reason. Everyone stated that the more important information that should be described to them is, in order of importance: who is talking, what is the action, the facial expression of the actors and the description of relevant specific objects.

4.3 Dataset

Four videos were shown to all participants. 1) A short film telling the story of a man meeting a woman on a beach (film 1). 2) A film excerpt from an action movie where a father and his son who are police officers and are in a conflicted relationship, have to investigate the same case (film 2). 3) An excerpt of a TV drama about the life of three doormen of a Montreal’s nightclub (drama). 4) A scientific report of a Canadian TV magazine (science) on the endangered tamarin monkey specie.

Film 1 had 4 main actors with few secondary actors in one scene in a public place. It contained a long scene without speech (73% of audio track) and was almost without background noise except music. In that scene, action was happening and we expected that it would be difficult for blind people without residual vision to imagine the action without asking for information. Film 2 (action movie) had many main actors and many secondary actors. The audio was mainly speech and noise (88% of the audio track) related to the on-going action.

Table 2. Description of dataset

Items	Film 1	Film 2	Drama	Science
Length (mm:ss)	08:49	06:12	07:35	06:49
Nb. Scenes	4	6	5	3
Nb. Actors	4	9	6	2
Secondary actors	Few	Many	Many	None
Nb. Speech Units	7	11	10	7
% Speech in Video	27%	88%	71%	91%
Nb. Non-speech Units	10	10	14	0
% Non-speech Units	73%	12%	29%	9%
Good contrast	Yes	Yes	No	Yes
Clear speech	Yes	Most	No	Yes

The non-speech segments were short and rare (12% of the audio). The drama had three main actors, four secondary actors and two scenes with a crowd. All the scenes had many dialogues (71% of the audio track) often set in very noisy environments. The science report was entirely based on the commentary of the narrator or the invited biologist (91% of the audio track). The visual content was complementing the discussion which to some degree did not need to be described.

5. RESULTS

Three levels of analysis are presented: 1) the analysis of requests made by Groups A and B classified as either confirmation or inquiry, 2) the distribution of these requests among the different information categories and 3) the comparison between the concepts found by the control group against the concepts stated by each participant. Our goal was not to compare Group A to Group B but to identify, facing the same data, 1) if different behaviors were adopted and 2) in which conditions, if any, they would adopt the same behavior.

5.1 Requests: Confirmation versus inquiries

Group A made more requests than Group B. They made 57.4% of the requests (253 out the 441 requests observed) while Group B made 42.6% of the requests (188 out of 441). As shown in Figure 1, Group A notably made more requests than B while viewing the films (19.0%, 18.1%). For film 1, the long silent scenes triggered many more from Group A, where B could follow the actors and the action since the contrast was good and there were not many people in the scenes. As for film 2, residual vision assisted Group B in identify the many actors and follow the numerous dialogues as opposed to Group A where they needed more cues to follow the dialogues. The science report generated fewer requests in both groups where Group A made slightly more requests than Group B (6.3% versus 4.8%). The lower number of requests is explained by the nature of the content which is mostly based on the narration and rarely on visual content. The results for the drama reveal an interesting fact. In this case, Group B made almost the same percentage requests (14.7% versus 13.8%) as in Group A. The drama had strong background noise which was a handicap for both groups since it contained conflicting auditory messages. Furthermore, the drama was composed of mostly night scenes with very low contrast. This took away the advantage of residual vision and made the need of information of Group B equivalent to the need of Group A.

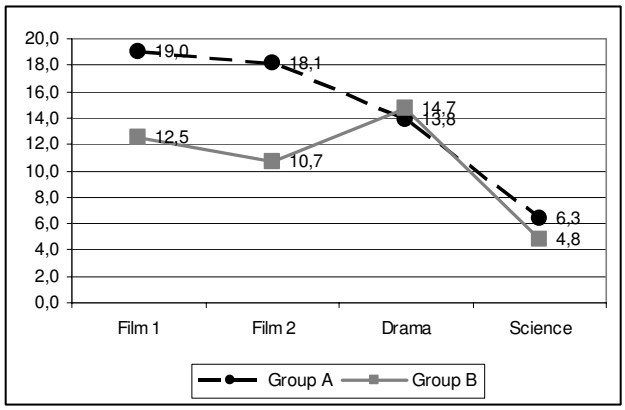

Figure 1. Percentage of requests per group

We analyzed the type of requests whether it was a confirmation or an inquiry. We considered a confirmation an indication of a lesser need for information since the person has prior knowledge but needs reassurance to avoid confusion. Contrarily to an inquiry which would indicate the need for an information to understand the meaning of the content.

Table 3. Percentage of type of requests

Data	% to confirm	% to inquire
Group A		
Film 1	39.3	60.7
Film 2	56.3	43.8
Drama	55.7	44.3
Science	60.7	39.3
TOTAL	**51.0**	**48.6**
Group B		
Film 1	67.3	32.7
Film 2	46.8	53.2
Drama	49.2	50.8
Science	33.3	66.7
TOTAL	**52.1**	**47.9**

From this analysis, we found that Group A made slightly more confirmations with 51.0% (129 out of 253) than inquiries with 48.6 (124 out of 253). The same behavior was observed in Group B with slightly more confirmations of 52.1 (98 out of 188) against inquiries at 47.9% (90 out of 188). However, results per data (Table 3) reveal changes in behavior depending on the content. Group A made more confirmations than inquiries for most content (Film 2, Drama, Science) except during Film 1 (39.3% versus 60.7%). This was caused by the long scenes without dialogue prompting the need for more inquiries.

The pattern observed in Group B is reversed. Except for Film 1, where they confirmed a lot more (67.3% versus 32.7%), Group B did more inquiries than confirmations. It is noticeable, for Film 1, how Group B confirmed a lot more while Group A inquired a lot more. Since Film 1 favored participants with residual vision, they gain more knowledge about the content. Moreover, the results for Film 2 and the Drama (both more complex data), induce the same behavior in Group A while in Group B, a change is observed. Although, for these two contents, they inquired more than confirmed (adopting a reversed behavior than GroupA), in the case of the drama, they confirmed almost as much as they inquired (49.2% and 50.8%). Again, we can see that the lack of contrast has the effect to reduce the advantage of residual vision and the behavior of Group B becomes similar to Group A.

The unexpected result was for the science report, where Group B made noticeably more inquiries (66.7%) than confirmations (33.3%) where we expected the inquiries to be lower than the Film 1. Our analysis of the distribution of requests among the different information categories revealed yet another interesting behavior.

5.2 Categories of Request

This analysis classified the requests made by participants into six categories: 1) who is speaking, 2) where the action is, 3) what is the action, 4) description of person or object, 5) the facial expression and 6) on the nature of sound. Our results showed that Group A made more requests on almost all categories except for the description (Figure 2).

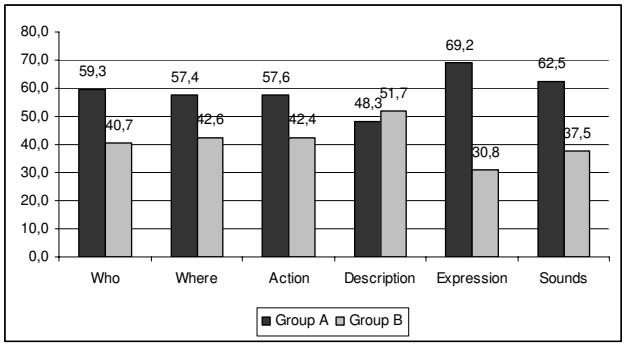

Figure 2. Percentage of category of requests

Their biggest requests were on facial expression (69.2%) and sounds (62.5%). Most requests on facial expression were to seek feedback on a dialogue when no verbal response could give this indication, for example: “He is happy when she asks him to follow her?”. Requests on sounds were either to know or confirm a specific noise (“What is that click?”) or to know if it conveyed meaning (“The music is playful, are the monkeys playing?”). Group B made slightly more requests about description (51.7%). Since their residual vision could help them identify information about the other categories, then they wanted more information on details they could not perceive clearly. For example, “What did he throw on the beach?” was asked when a men put down his jacket before sitting down.

In Table 4, a more detailed analysis per film shows the percentage of requests made per category. The most frequent requests for all contents were about “action” and for both groups. The highest scores being for Film 1 and Science in the two groups (Group B 60.0% for Film 1 and 57.1% for Science while in Group A: Science 53.6% and Film 1 52.4%). A more detailed analysis of the requests on action revealed that for Group A, 68.2% of these were inquiries and in Group B, 66.7% were confirmation of action. This indicates that knowing the action is crucial to the understanding of the film and that even with residual vision this information is more important than to know who is in the scene. This is even more evident when we look at the results of simple content like Film 1 and Science (few actors and good contrast) that have low percentages for those categories (in Group B, there were no request on “who” for the Science).

Table 4. Percentage per Category in Requests

Category	Film 1	Film 2	Drama	Science
Group A				
Who	11.9	30.0	26.2	3.6
Where	11.9	13.8	13.1	7.1
Action	52.4	40.0	37.7	53.6
Description	7.1	7.5	11.5	32.1
Expression	8.3	-	3.3	-
Sound	8.3	8.8	8.2	3.6
Group B				
Who	14.5	25.5	23.1	-
Where	7.3	12.8	13.8	19.0
Action	60.0	40.4	41.5	57.1
Description	9.1	10.6	12.3	-
Expression	7.3	-	-	-
Sound	1.8	10.6	9.2	-

The requests made on “sound”, “expression” and “description” categories are similar for all contents (except for “description” in Science by Group A). This seems to indicate a basic need for these categories that can be even omitted in certain cases, i.e. when the content is self-explanatory (Science) or the complexity is already overwhelming (Film 2 and Drama).

Also, noisy background had a stronger effect on Group B since the requests by Group A were more constant for all contents (8.3 for Film 1, 8.8 for Film 2 and 8.2 for Drama) with a slight decrease for Science (3.6). However, we observed in Group B that those requests increased noticeably during Film 2 (10.2) and Drama (9.2), caused by the noisy background of both contents with conflicting auditory channel.

The previously noted that exception on description for Science was induced by one participant in Group A, who stated to have a very strong interest for report on animal and asked many detailed questions that were not raised by any other participants. For instance: “The monkeys had how many fingers? “Are there many trees in the environment and are they all tropical?”. We considered his percentage an outlier but kept it in the results because we could not discriminate fairly which requests could be considered as not “standard”.

In the earlier analysis of confirmation versus inquiry, we found an unexpected result where Group B made more inquiries than confirmations for the science report. We observed that there requests increased for two categories: action and where. The

requests on action were mainly to obtain more detail about the action of the monkeys which had dark furs on a dark green background. So, most of them saw the two monkeys moving but could not perceive clearly their action. This relates again to the effect of bad contrast on the needs of Group B. As for the place (where), some of them were confused with the images of what seems like a jungle and the mention in the narration that this was taking place in the Biodôme. Since the narration did not explain the natural habitat provided in the Biodôme but they could perceive a jungle, this was a source of confusion.

Finally, the category analysis made on complex contents (Film 2 and the Drama) reveals that the needs for certain categories are crucial for some contents. Furthermore, it also stimulated the adoption of same behavior by both groups. Complex contents generated by the highest requests are on "actions", "who" and "where". The larger number of actors could create more confusion, especially if there are more dialogues. Many actors, many dialogues and many scenes are difficult situations for participants in both groups. And again, the bad contrast of some scenes was a handicap to Group B. This further indicates that when VII are faced with complex contents where any residual vision advantage is lost then the quantity of information and the category of information needed will be the same for individuals of levels 5, 4 and 3 of WHO's classification. Overall, those results indicates that Group B adopted a behavior similar to Group A when confronted to a complex content such as too many actors, confusing visual information, bad contrast and conflicted auditory channel. These situations do not provide them with discriminating visual elements they needed.

5.3 Mental representations

Our last analysis compared the concepts recalled by the participants versus concepts identified by a sighted control group. The control group had identified 12 concepts for the Film 1, 9 for the Film 2, 14 for the Drama and 5 for the Science report (Table 5). The lower results for both groups were for Film 1 with 8.2 on 12 and Drama with 8.1 on 14. Both contents were already identified as the more complex ones in the dataset which could explain the lesser number of concepts retained by the VII.

The results revealed that Group B omitted on average more concepts than Group A. These results do not indicate that Group A had a better understanding than Group B. This would be a hasty interpretation of the data, since we found that most omitted concepts were stated in earlier analysis. A correlation of this data with the confirmed requests done by Group B indicated that most of them were requested as a confirmation. This indicates that Group B knew about the concepts but choose not to include them in their summary. More detailed interviews would need to be conducted to understand this observation.

Table 5. Average of found concepts.

Data	Control group	Both groups	Group A	Group B
	Total nb. concepts	Average	Average	Average
Film 1	12	8.2	9.0	7.4
Film 2	9	6.4	6.6	6.2
Drama	14	8.1	8.6	7.6
Science	5	3.7	3.6	3.8

Furthermore, the fact that Group A stated on average more concepts that Group B could also be partially explained by their better memory capacities. Indeed, as mentioned earlier (Sadato, 2002), there is evidence suggesting that congenital blinds and potentially the late blinds (before the age of 16) could store more information in their memory than late blind after the age of 16. Since, four of the five persons in Group A meet this criterion than these individuals would have been able to recall more concepts.

6. DISCUSSION

Our goal was to study how VII seek information from their relatives in a real-life situation when they are watching audio-visual content without VD. We aimed at gaining a comprehensive understanding of their abilities and limitations in order to customize VD to their needs. Our approach which was to analyze the verbal protocol of the participants while listening to audio-visual content with a sighted assistant, proved to be fruitful in provide insights on how we can better render VD. So their need may not be about getting less VD, but a need for VD to provide auditory information that requires less attention and thus changing an ordeal into an entertainment, and avoid a cognitive overload.

The basic quantity of VD needed by VII with residual vision (level 3) would be less than the quantity required by VII in levels 1 and 2 for contents that are simple. These contents are the ones with few actors, with no conflicted auditory channel and with good contrast. In a very simple example such as a science report with one clear narrating voice, the VD would be minimal and would easily address the needs of individuals with and without residual vision. In another example as a simple film with few actors, VII without residual vision would need more VD than individuals with residual vision. So, different degrees of VD should be made accessible on demand. Moreover, if the need for confirmation arises there should be a mechanism for them to validate the information without every detail being in the VD.

Furthermore, we found that the advantages of individuals with residual vision can be quickly overturned when viewing conditions are not optimal. Complex contents have many actors, with many dialogues with potential conflicted auditory channel, possibly many scenes and bad contrasts. Any combination of these constraints should be made visible to the VD scripter so that he can produce a VD that limits these handicap conditions. In those cases, VII with or without residual vision will need the same basic amount of information. However, additional VD should be given to counter the effects of handicap conditions in order to better discriminate among the actors, to inform on the change of scene and to provide details on scene with bad contrast. In this context, a mechanism for confirmation would be even more helpful since they may be very few places to add VD.

Based on those results, we concluded that we needed implement and test a VD player that could offer those features to the VII. These tests would validate the proper level of information needed and the capacity of the technology to make VII autonomous when they consult audio-visual contents. The implemented player does not only provide a basic quantity of information but also offers the possibility to give more VD if needed, and the ability to confirm information on demand. The basic information is the VD that can be entered into the available gaps in dialogues without extending the duration of the contents. The VD is scripted to provide the category information in the order of importance found in this study. Also, any handicap condition that can be treated in the time allocated is also included. The extended level offers more VD to

complete the information that could be needed by VII without residual vision and also all encounter handicap conditions.

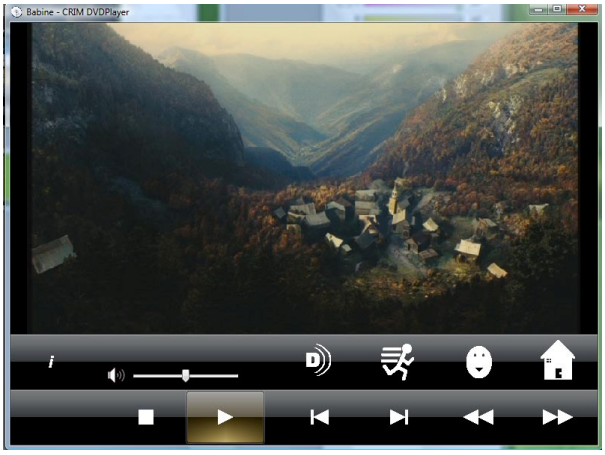

Figure 3. CRIM's VD player

As seen on figure 3, our VD player integrated a choice of VD with the "D))" button that offers a toggle mode between a standard and an extended version of VD. We also implemented three on-demand information buttons: 1) to state all the actions of the current scene, 2) to name the actors of the current scene and 3) the name of the scene.

Our next step is to conduct other interviews with individuals with low vision that are classified levels 1 and 2 by the WHO. Our aim is to gain a comprehensive understanding of the cognitive and memory capacities of a large spectrum of individuals with visual impairments to assess their needs for VD and to design tools that are truly accessible.

Furthermore, we are planning to test an experimental VD service that would provide the expected 30 participants with the VD player and VD production implemented from the findings of the in-situ study.

7. ACKNOWLEDGMENTS

Our thanks to Ms Anne Jarry, professor in Visual Rehabilitation at the School of Optometry of the University of Montreal, for her helpful suggestions and unlimited expertise. We are also very grateful to the participants who so generously gave their time and insights to make this research possible.

8. REFERENCES

[1] World Health Organization (2010), http://www3.who.int/currentversion/fr-icd.htm

[2] Chapdelaine, C., Jarry A. (2010). Lessons learned from Blind individuals on Videodescription, Proceedings of AHFE, Miami.

[3] Mergier, J. (1999) Le classement OMS des déficiences visuelles, http://www.irrp.asso.fr/articles/article007.html

[4] Presley, I., D'Andrea, F.M. (2009) Assistive Technology for Students Who Are Blind or Visually Impaired: A Guide to Assessment, AFB Press, 500 pp.

[5] Ponchillia P.E., Ponchillia, S.V. (1996) Foundations of Rehabilitation Teaching with Persons Who Are Blind or Visually Impaired, AFB Press, 432 pp.

[6] Sadato, N., Okada, T., Honda, M., and Yonekura, Y. (2002) Critical period for cross-modal plasticity in blind humans: A functional MRI study. Neuroimage, 16, 389-400.

[7] Colavita, F.B. (1974) Human sensory dominance. Perception and Psychophysics, 16, 409-412.

[8] Wickens, C.D., Holland J.G. (2000) Engineering Psychology and Human Performance, 3rd Ed. Upper Saddle River, NJ, Prentice-Hall, 572 pp.

[9] Gouzman, R. and Kozulin A. (2000) Enhancing Cognitive Skills in Blind Learners. The Educator: 20-29.

[10] Reisner, J. (2008) Theory and Issues in Research on Blindness and Brain Plasticity. In Blindness and Brain Plasticity in Navigation and Object Perception, Edited by Reiner, Ashmed, Ebner and Corn, Lawrence Erlbaum Associated 423pp.

[11] Merabet, L.B., Bass Pitskel, N., Amedi, A., Pascual-Leone, A. (2008). The Plastic Human Brain in Blind Individuals: the Cause of Disability and the Opportunity for Rehabilitation. In Blindness and Brain Plasticity in Navigation and Object Perception, Edited by Reiner, Ashmed, Ebner and Corn, Lawrence Erlbaum Associated, 423pp.

[12] Vanlierde, A., Renier, L. De Volder, A.G. (2008) Brain Plasticity and Multi-Sensory Experience, In Blindness and Brain Plasticity in Navigation and Object Perception, Edited by Reiner, Ashmed, Ebner and Corn, Lawrence Erlbaum Associated, 423pp.

[13] Rokem, A. and Ahissar M. (2008) Interactions of cognitive and auditory abilities in congenitally blind individuals. Neurophychologia, 47, 843-848.

[14] Gagnon L., Foucher S., Héritier M., Lalonde M., Byrns D., Chapdelaine C., Turner J., Mathieu S., Laurendeau D., Nguyen N.T., Ouellet D. (2009) Towards Computer-Vision Software Tools to Increase Production and Accessibility of Video Description to Visually-Impaired People, Universal Access in the Information Society, Springer-Verlag, Vol. 8, no. 3, 199-218.

[15] Gagnon L., Chapdelaine, C., Byrns, D., Foucher, S., Héritier, M., Gupta, V. (2010) Computer-Assisted System for Videodescription Scripting, Proceedings of Computer Vision Application for Visually-Impaired (CVAVI), a satellite workshop of CVPR 2010, San Francisco, (to be published).

[16] Chapdelaine, C., Gagnon, L. (2009). Accessible Videodescription On-Demand. In Eleventh International ACM SIGACCESS (ASSETS'09). Pittsburgh, PA, USA, October 26-28.

[17] Ofcom (2000). ITC Guidance on Standards for Audiodescription. http://www.ofcom.org.uk/tv/ifi/guidance/tv_access_serv/archive/audio_description_stnds.

[18] ADS (2009). Guidelines for audiodescription (initial draft of May 2009). http://www.adinternational.org/ad.html

[19] Morisset L., Gonant F. (2008). Charte de l'audiodescription. http://www.travail-solidarite.gouv.fr/IMG/pdf/Charte_de_l_audiodescription_300908.pdf.

[20] Braun, Sabine (2008). Audiodescription research: state of the art and beyond. Translation Studies in the New Millennium, Vol. 6, 14-30.

[21] Turner, J. (1998). Some Characteristics of Audio Description and the Corresponding Moving Image. Proceedings of 61st ASIS Annual Meeting, vol. 35, 108-117. Medford, NJ: Information Today.

[22] Turner, J., and Mathieu S. (2008). Audio description text for indexing films. International Cataloguing and Bibliographic Control 37, no. 3 (July/September), 52-56.

[23] Piety, P. (2004). The language system of audio description: an investigation as a discursive process. JVIB 98:8. 453-469.

[24] Salway, Andrew. 2007. "A Corpus-based analysis of the language of Audio Description". In Media for All, Díaz Cintas, Jorge, Pilar Orero and Aline Remael, eds. 151-174.

[25] Peli E., Fine E. and Labianca A. (1996). Evaluating visual information provided by audio description. JVIB 90:5. 378-385.

[26] Pettitt B., Sharpe K. and Cooper S. (1996). AUDETEL: Enhancing television for visually impaired people. BJVI 14:2. 48-52.

[27] Schmeidler E. and Kirchner C. (2001). Adding audio-description: does it make a difference? JVIB 95:4. 197-212.

[28] Ely R., Emerson R. W., Maggiore T., O'Connell T., & Hudson L. (2006). Increased content knowledge of students with visual impairments as a result of extended descriptions. Journal of Special Education Technology, 21(3), 31-43.

[29] Suchman, L. (1987). Plans and Situated Actions: The Problem of Human-machine Communication, Cambridge University Press, New York.

[30] Roskos-Ewoldsen B., Roskos-Ewoldsen D. R. and Yang M., (2003). Testing the Landscape Model of text comprehension. Paper presented at the annual meeting of the International Communication Association, San Diego, CA.