View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by AIS Electronic Library (AISeL)

Association for Information Systems
AIS Electronic Library (AISeL)

AMCIS 2006 Proceedings

Americas Conference on Information Systems (AMCIS)

December 2006

LiveDescribe

Carmen Branje
Ryerson Univerity

Susan Marshall
Ryerson Univerity

Ashley Tyndall
Ryerson Univerity

Deborah Fels
Ryerson Univerity

Follow this and additional works at: http://aisel.aisnet.org/amcis2006

Recommended Citation

Branje, Carmen; Marshall, Susan; Tyndall, Ashley; and Fels, Deborah, "LiveDescribe" (2006). *AMCIS 2006 Proceedings*. 368.
http://aisel.aisnet.org/amcis2006/368

This material is brought to you by the Americas Conference on Information Systems (AMCIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in AMCIS 2006 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

LiveDescribe

Carmen Branje
Ryerson University
cbranje@ryerson.ca

Susan Marshall
Ryerson University
smarshal@ryerson.ca

Ashley Tyndall
Ryerson University
atyndall@ryerson.ca

Deborah I. Fels
Ryerson University
dfels@ryerson.ca

ABSTRACT

Television, film and digital media are important cultural experiences that help shape and define society. Exclusion from these experiences may present a barrier to wider participation in society. While video description of pre-recorded content (e.g., movies, many TV programs, etc.) is becoming more common, video description of live programming (e.g., news, improvisational theatre, etc.) has received little attention and very few examples are available to illustrate its feasibility. We present a tool, LiveDescribe, which enables the production of real-time video description for on-line broadcasts. LiveDescribe includes features such as two-button operation, and automatic detection of silent periods that enable it to be used by a single operator. Initial feedback from describers about the usability of LiveDescribe is relatively positive but they found the task of live description difficult. The cognitive load imposed by the real-time aspect of describing tasks is high and may lead to high error rates.

Keywords

Live video description, live audio description, accessible media

INTRODUCTION

Television, film and digital media are important cultural experiences that help to shape and define society. Being able to share and participate in these experiences on an equal footing is a crucial element of social and cultural accessibility for people with disabilities.

Access to video content (on-line and television) for people with vision impairments can be achieved by providing descriptive video information (DVI), also known as "described video™" and "audio description", a technology that has only recently received considerable attention. Using the secondary audio program (SAP) channel on most newer televisions, DVI provides spoken descriptions of visual content of a video, inserted into "usable" silent or quiet periods in the original soundtrack (CRTC, 2004).

Schmeidler & Kirchner (2001) showed that DVI significantly improves comprehension of audio-visual presentations, particularly in cases where information is presented solely in a visual manner, with no accompanying dialogue or explanation (e.g., nature programs). Their study also highlights other benefits of DVI: users becoming more comfortable speaking with sighted people about the same program, not having to rely on informal descriptions provided by friends or family, and allowing users to participate in shared cultural experiences on a more equal level.

The process of creating video descriptions for pre-recorded television broadcasts or movies is similar to the production process of the video content itself – a script is written, a narrator is hired and the video descriptions are read and edited (Office of Communication, 2004). For live broadcasts or webcasts, however, there is no standard production process for real-time descriptions and very few examples (Media Access Group, 2004).

In on-line environments, post-production described video may be produced using the same model used for pre-recorded television broadcasts. However, very little described video is offered on-line and only a few guidelines and recommendation documents are available. The Web Content Accessibility Guidelines (WCAG) recommends having video description for on-line video content (W3C, 2003). However, it appears that considerations are being made to move this from checkpoint 1.2 to a best practices section. This may be due to the popularity of webcams, live webcasts, and the difficulty and lack of process for providing live descriptions.

There are several examples of live description that do provide some useful insights. These include describing services at some theatres (Weeks, 2002) and the long-established practice of live sporting event commentaries. In most cases, live theatre description is offered by a well-trained volunteer who spends time preparing before the performance by attending dress rehearsals or early performances, and reading the script (Audio Description Associates, 2004). In the case of live sporting events, play-by-play commentary involves an experienced commentator relating the action as it is happening. The information conveyed by this commentary is usually far more useful for people who have experience with the game than those who have never seen it played, or had it properly explained.

Live description is likely prone to higher error rates than post-production description, because it is delivered live and cannot be reviewed or edited. However, there is little research on rates of error for theatre describers (or what constitutes adequate description), and a lack of quantifiable acceptance ratings for DVI quality by people who are blind or low vision.

We present a new software application, LiveDescribe, which has been developed to allow a single individual to analyze content and provide video description in near real-time. We also present preliminary describer reactions to and workload measures of a live describing task, and responses to using the LiveDescribe tool for the process.

THE LIVEDESCRIBE SYSTEM

LiveDescribe is designed to allow near real-time DVI to be added to on-line content by DVI describers (the target user group for LiveDescribe). While there are a number of existing software applications such as Softel ADePT (Softtel-USA, 2001) and Magpie (NCAM, 2003) that support post production DVI, they are not designed to support live description for events such as emergency broadcasts, parades, concerts, and other events which are often viewed only once, and relevant only for the time in which they appear.

The real-time description process requires a different approach from post-production description. A real-time describer must record descriptive utterances based on a much shorter duration and constrained view of the material. There is no opportunity to fully preview the material before a description is voiced. In order to support this type of description, LiveDescribe provides an automated service that:

1. determines the amount of time before the next describable period (i.e. time intervals during which sound is non-dialogue or paralinguistic, and is low volume). A describer can choose to create a description during these periods in real time (see Figure 1). For LiveDescribe, an entropy algorithm is used to differentiate between dialogue and non-dialogue portions of an audio signal. As seen in Figure 1, an entropy graph of the audio signal along with a threshold line is displayed in the user interface so that users can easily see when dialogue and non-dialogue periods occur. A non-dialogue period is defined by the portion of the entropy graph that falls below the threshold line and between the peaks of the neighbouring portion of the graph that is above the line.
2. allows descriptions to extend beyond the limits of silent periods. The video automatically pauses (extending its running length) to accommodate descriptions that exceed the detected periods of silence. Professional describers have noted the importance of this feature to provide richer descriptions. It also allows describers to be less accurate with the description timing and the silent periods, which is important when there is little opportunity to plan the description content. However, video extension cannot be used when there are external time constraints for presentation, such as in broadcast television.

Despite these automated tools, we predict a live description task will impose a high cognitive demand on a describer. The describer must attend to two main activities while the video content is playing: 1) the occurrence and length of silence periods, indicated by the silent period indicator; and 2) the content that must be described. The describer needs to decide what is important to describe for that particular scene, determine the most appropriate aspects of a scene to describe, and finally insert the description as the video plays. In addition, a describer does not have an innate ability to predict which aspects of the scene that he is describing will be essential to the overall story until later scenes. However, if a describer has experience with, or domain knowledge for a specific live show, or even the genre of that show, decisions about what it important to describe can be easier.

Ease-of-use is critical in supporting describer activities because of this high cognitive and task load. We attempt to reduce the cognitive load imposed through the user interface by using a single button control for the entire process. This button is pressed once to start describing and pressed again to stop. A hands-free function such as a foot pedal or voice activation may also reduce the attention required for the interface and could be easily implemented if describers found it desirable. We have also added a graphical countdown timer and time remaining display that inform describers of how much time is left for the remaining description of that particular describing opportunity.

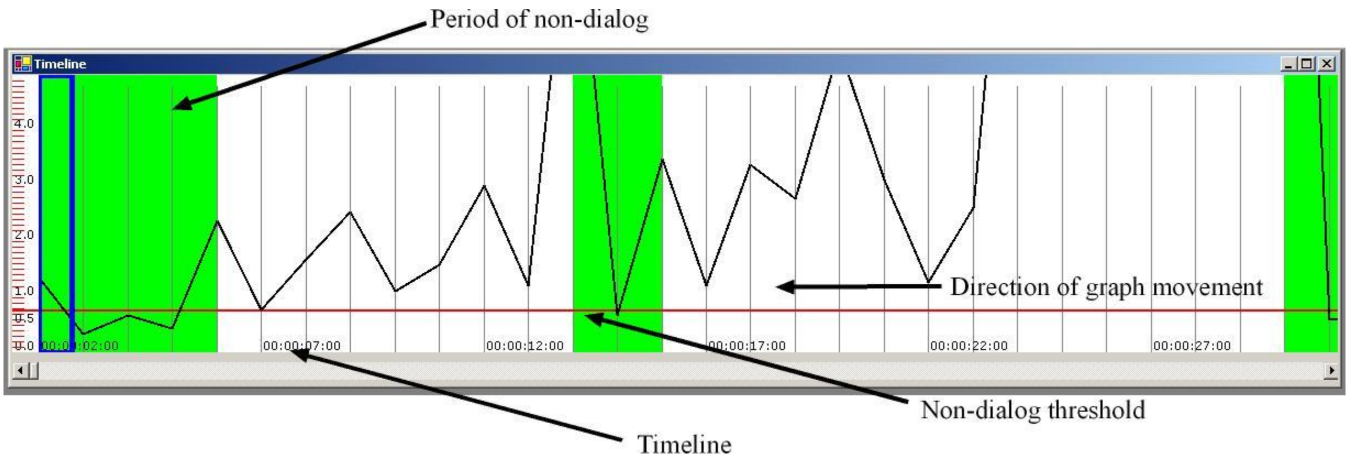

**Figure 1: Graph of a sample audio track from a television show.
Non-dialogue periods (length and occurrence) are shown as coloured rectangles.**

The system monitors the sound level in the source video at all times. If sound continues after a silent period and the describer has not stopped describing, the source video is automatically paused and extended description can continue. Once the describer stops describing, the video resumes. One drawback of this pause during extended description is that it may be confusing for sighted viewers.

While some of these interface features are similar to those found in the Softel's ADePT suite (e.g., look-ahead bar, a countdown timer, time remaining displays, and one touch recording; Softel-USA, 2001), live description, sound level monitoring and extended description functionality are unique to LiveDescribe.

PROCEDURE FOR DESCRIBING A LIVE STREAM

The procedure for describing a live on-line video stream involves three main steps. First, the describer must identify the video source and load/stream it to the source video window (see Figure 2). Next, the user must determine how much "look ahead" she requires. Look ahead times can be selected between thirty seconds and five minutes. This then defines the quantity or portion of video material to be processed at one time. When determining the look ahead time, a describer must make an important tradeoff, for, although a longer the look ahead time gives her more time to preview the video material, it also means that the processing time becomes greater, causing increasing delays in delivering the video material to the audience. Once the look ahead time is set by the user, the system then looks ahead, stores and analyzes that portion/quantity of the video for non-dialogue periods. This process causes the described version of the presentation to be delayed by twice the length of the defined look-ahead value because the video cannot be buffered and analyzed for non-dialogue periods simultaneously. The clip must first be recorded fully (one look ahead time period), and then analyzed for non-dialogue periods (analysis takes approximately one look ahead time period) before it can be described. Only then can the next portion of the video be buffered. The system is thus only near to real-time because of the delay introduced by the look-ahead and analyze times. However, it provides the describer with at least 2- minutes of advanced notice and some opportunity to prepare for upcoming silent periods. Whether or not this is enough preparation time depends on the visual complexity of the content, describer knowledge of the content style and the frequency of periods appropriate for description.

The second step involves determining which silent periods to use for description. The LiveDescribe interface displays the number and length of the non-dialogue periods within the defined portion. This is displayed as a rectangular bar located along the graph, called the space highlighting bar as seen in Figure 2. The width of each bar indicates the length of time of each period. As time progresses, the solid rectangular bar will decrease in width indicating the time elapsed and the number of seconds remaining in the usable describing period.

The describer's last step is to record descriptions for the video material. The describer accomplishes this by pressing the "Record" button (a red circle) and speaking into a recording microphone. Once the describer finishes her description, she presses the "Stop" button (a grey square). Status indicators next to the record button indicate the description status (being recorded or not) as well as whether the content is "auto extending". This causes the video channel to pause while the audio channel continues broadcasting the description as the describer speaks it.

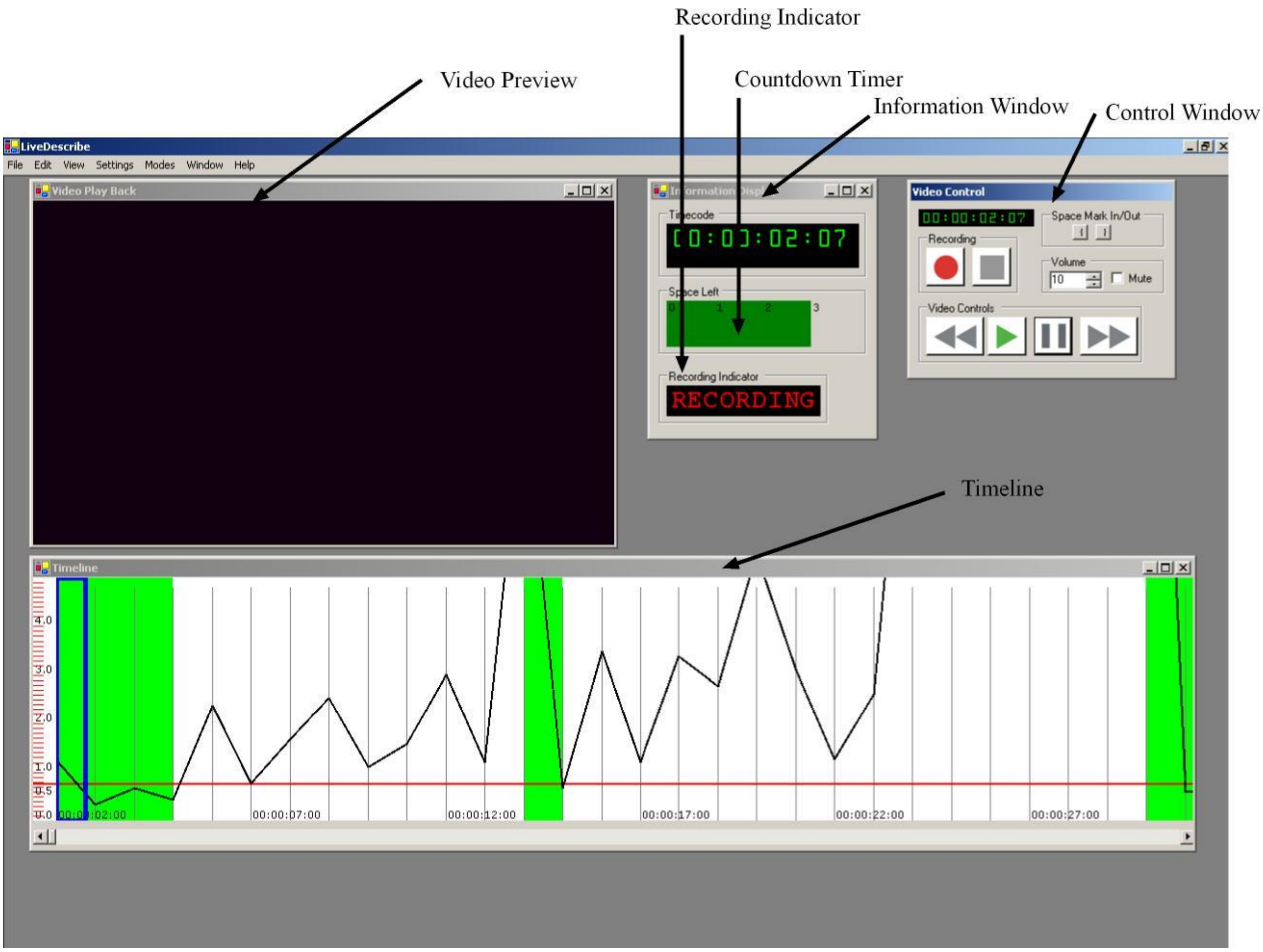

Figure 2. LiveDescribe Interface

Preferences

It is important to provide describers with control over the description recording environment, particularly since the task of real-time describing is likely to be very difficult. There are several user selectable preferences that a describer can change in order to make the live description environment to accommodate his needs and/or preferences.

The "auto pause" feature can be toggled on or off as certain content (e.g., live television broadcasts) may not warrant this type of description. The look-ahead time is selected by the describer and this selection will be dependent on describer comfort levels (longer time allows describers to see more content before describing), number of expected non-dialogue spaces, and the length of those spaces.

The describer can also define the length of a non-dialogue period that would be considered as acceptable describing opportunities by the system. For example, a user may want to ignore all non-dialogue periods shorter than 2 seconds, as this does not leave enough time for a valuable description. The colour of the dialogue indicator bar can also be changed to suit the preferences of the user.

User Feedback

Audio describers are the current target user group for evaluating the LiveDescribe tool. Viewers who are blind or low vision are the eventual users of resulting descriptions produced by describers using LiveDescribe. However, we must first evaluate the feasibility and usability of LiveDescribe with describers before proceeding to outcome testing with viewers who are blind or low vision. In this paper we report questionnaire and NASA Task Load Index (TLX) subjective workload data gained from our initial subjects.

Method

We selected a soap opera, The Young and the Restless, as the simulated live content. As soap operas are broadcast daily, there is a daily production schedule which does not leave much time for descriptions to be produced using the post production method. In addition, domain knowledge can be easily acquired. Two days before participating in the study, participants were provided with the previous episode of this show in order to become oriented to the next episode which was used in the study, and to become familiar with the characters, scenes, and props in the show.

Two male describers participated in the initial pilot study of LiveDescribe. One describer had over eight years experience as a describer and the other had four years experience describing television and film. In a pre-questionnaire administered at the beginning of the actual study, subjects were asked about their experience with conventional video description and to outline the techniques they used to produce descriptions. They were also asked whether they had live description experience and to speculate on whether it would be feasible.

The average amount of time reported to complete a video description script for a one hour television show was seven hours. The script writing process that both describers used consisted of the following steps: 1) view the show first; 2) create a first draft of the description script by reviewing the show scene by scene; 3) edit the script by reading it back to the video to check that it is accurate, objective, consistent, and fits with the timing of the images and dialogue so that there is no overlap with the dialogue.

Neither describer had any experience with describing live events. However, both describers thought that the task of live describing might be feasible but that correct timing and prior subject knowledge would be important for success. The more experienced describer suggested that not having time to reflect on the description could cause errors and significantly reduce description quality.

For the study, participants were fitted with a heart rate monitor and asked to relax until a steady state resting heart rate was established. Participants were then trained in using the LiveDescribe software and practiced using unrelated video content, a convocation video. Participants were then asked to watch and video describe the soap opera in three 12-minute sections to simulate commercial breaks and to allow the describer to have a short break. Following the describing session, participants were asked to complete a NASA TLX subjective workload survey (Hart and Staveland, 1988), and a post-study questionnaire. For the NASA TLX workload assessment participants provide two measures: 1) a weighting or level of importance of six validated workload factors (mental demand, physical demand, temporal demand, performance, effort and frustration - definitions for each factor can be found in Hart and Staveland (1998); and 2) a low/high rating from 0 to 100 (where 0 is a low and 100 is a high amount) of the magnitude of the effort required in each of those factors. The combination of weighting and rating provides the workload scores for each factor and a total for the task. The post study questionnaire collected subjective reaction to the process of live describing, level of difficulty of live describing tasks, fatigue levels, error rates, and speculate about the response of blind viewers to their live description.

RESULTS

Only one subject was able to complete the NASA TLX rating for the study and thus only one data set is reported here. However, it does provide some insight into what factors are considered important to a describers workload for live describing tasks. Figure 3 shows the combined workload score (weight times rating) for each workload factor. The describer believed that the frustration factor contributed most (33%) to workload followed by effort (20%). Mental demand, temporal demand and performance contributed equally (13% each), and physical demand contributed the least (7%). However, the participant rated the magnitude of required effort fairly moderate at 50 out of 100 whereas performance and frustration received highest rating of 100. This indicates that while the participant believed that effort was important to consider in determining workload, he believed that he only had to work moderately hard to accomplish his level of performance.

This subject weighted frustration as contributing the most to his sense of workload and also rated his level of frustration with the task at the highest level for an overall workload score for this factor of 33.33. The participant also rated his success in accomplishing his task (performance) as very poor (the worst possible rating at 100), however, the weight attributed to the performance factor was only 13% for an overall workload score for performance of 13.3. The participant rated mental demand (defined as the quantity of thinking, decision-making, searching and remembering) as fairly high (70 of 100) while only weighting this factor at 13% contribution to workload. However, the overall workload score for this factor is fairly high at 9.33 compared with the remaining factors. Finally, the overall workload for this subject was 70.7 (out of 100) and is considered to be high. For comparison, France et al., (2005) found that the overall mental workload of emergency room physicians carrying out standard emergency room patient tasks was 71.8.

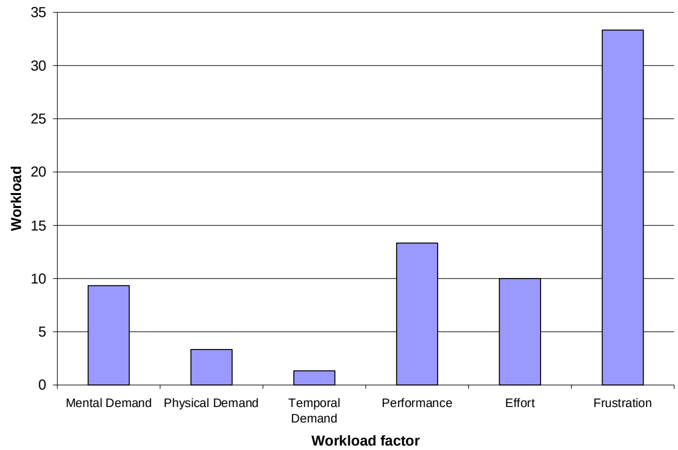

Figure 3: Overall cognitive workload for each workload factor

As this is preliminary data, no statistical analysis is warranted for the post-study questionnaire results. However, a summary of some of the important indicators is presented. When compared with the post-production process of video description, both participants agreed that it was very difficult, confirming the workload results. They also agreed that they made many more severe errors such as using incorrect names of people or locations, or talking over the dialogue. In addition, both participants reported feeling tired after the first 12-minute session or during the second 12-minute session. They diverged on the amount of time required before a break with one person suggesting it was acceptable between 5 and 15 minutes of live describing, and the second suggesting after one half hour of describing would be reasonable (in this question neither describer selected the option that description was not possible).

The reaction to the level of difficulty of using LiveDescribe software varied when participants were asked their opinion of various features. Specifically, the size and place of the video window was somewhat difficult but the "look ahead" and keyboard shortcuts were somewhat easy to use. One participant reported that the space highlighting was somewhat easy to use while the other thought it was somewhat difficult to use. Finally, one participant reported that the size and placement of the video window was easy to use and the other thought it was difficult to use. Both participants were concerned with the accuracy of the space highlighting and recommended that it be improved.

DISCUSSION AND RECOMMENDATIONS

Currently, there are no standards of practice and few examples of live description for television or on-line media. Much more experience and research is required in order to make adequate and comprehensive recommendations and best practice guidelines. However, initial reactions to LiveDescribe by describers have highlighted key areas for further work.

Similar to live captioning, we predict that live description will be more error prone, more stressful and more fatiguing than post-production description. The mental workload measures and the post-study responses regarding these factors initially support this hypothesis, although only limited data is available.

In a live describing situation there is no opportunity for review, writing scripts or making corrections, so any preparation must be carried out before the actual event occurs. Both subjects agreed that being familiar with the content type and style prior to a live description is critical. In this study, neither subject had any prior experience or knowledge with the particular soap opera that was used. However, they were provided with the previous days' episode to gain some experience. Both subjects suggested that this was sufficient preparation for them. However, the level of difficulty and cognitive workload experienced by these describers would like decrease as they gained more expertise with the content area. Longitudinal studies with similar content are recommended to investigate the impact of experience on the accuracy and fatigue rates of live description.

We also suggest that even though the live description process is difficult and stressful, it is still doable with content expertise. Both describers agreed that live description was possible but that it was error prone and stressful. One important challenge in designing live description processes and tools involves determining the quantity and types of errors (e.g., omission, mis-descriptions, etc.) produced by describers for a variety of live events and timeframes, and the level and quantity of errors that would be acceptable to viewers..

Technical Challenges

To facilitate live description, the necessary tools need to be made available to support the process. LiveDescription is one tool that attempts to do this. The initial usability results suggest that the “look ahead” and keyboard shortcuts of the tool are important and, in the current implementation, easy to use. Improvements in the accuracy of the non-dialogue spaces and the graphical display of that information are required. Towards this goal and as a result of this initial data, we are refining the non-dialogue detection algorithm by adding additional second-order processing algorithms such as zero-crossing and frequency response. The goal here is to achieve 100% accuracy in non-dialogue detection and determination of the describing space. At the moment the current algorithm is at about 80-90% accuracy.

We have developed a prototype tool that allows near live video description to be produced and have carried out limited evaluations. These limited evaluations have identified improvements to the tool as well as provided initial indicators of the level of difficulty and workload imposed by the live description process. There is still considerable research required to determine acceptable quality parameters, efficient methods and interface designs for describers, and ways to reduce the high cognitive load and possible high fatigue rates imposed on describers by live description tasks.

ACKNOWLEDGEMENTS

Funding for this project is provided by the Department of Canadian Heritage. We thank Mark Rosen and John Stubbs of Audio-Vision Canada for all of their advice and assistance in this project. The authors also wish to gratefully acknowledge the University of Toronto and Ryerson University.

REFERENCES

1. Audio Description Associates Inc. (2004). Training. Retrieved March 9, 2004 from http://www.audiodescribe.com/training1.html.
2. Canadian Radio and Television Commission. (2004). Broadcasting Public Notice CRTC 2004-2. Retrieved February 7, 2004, from http://www.crtc.gc.ca/archive/ENG/Notices/2004/pb2004-2.htm.
3. Clark, J. (2003). Building Accessible Websites. Boston: New Riders.
4. France, D. J., Levin, S., Hemphill, R., Chen, K., Rickard, D., Makowski, R., Jones, I. and Aronsky, D. (2005) *International Journal of Medical Informatics*, **74,** 827-837.
5. Hart, S. G. and Staveland, L. E. (1988) In *Human Mental Workload* (Eds, Hancock, P. A. and Meshkat, N.), pp. 139-183.
6. Media Access Group - WBGH. (2003). Real-time captioning -- a career with a meaningful mission. MAG Guide Vol. 11. Retrieved on March 26, 2004, from http://main.wgbh.org/wgbh/pages/mag/resources/guides/mag_guide_vol11.html.
7. Media Access Group - WGBH. (2004). Live Description for Live Programs and Teleconferences. Retrieved March 21, 2004, from http://main.wgbh.org/wgbh/pages/mag/services/description/.
8. National Captioning Institute. (2004). NCI’s Live Captioning. Retrieved March 10, 2004, from http://www.ncicap.org/preview/livecapproc.asp.
9. NCAM. (2003). MAGpie 2 Documentation : Contents. Retrieved on November 26, 2004, from http://ncam.wgbh.org/webaccess/magpie/magpie_help/
10. Office of Communication. (2004). Subtitling, Signing & Audio Description. Retrieved March 21, 2004, from http://www.ofcom.org.uk/codes_guidelines/broadcasting/tv/sub_sign_audio/?a=87101.
11. Schmeidler, E., Kirchner, C. (2001). Adding audio description: Does it make a difference? Journal of Visual Impairment & Blindness, 95, 197-203.
12. Softel-USA. (2001). ADePT Audio Description Technology. Retrieved November 26, 2004, from http://www.softel-usa.com/l4_products_subtitling_adept_main.php3
13. W3C. (2003). Web Accessibility Initiative. Retrieved October 3, 2003, from http://www.w3.org/WAI/.
14. Weeks, J. (2002). Bringing ballet to the blind. Dance Magazine. 23-24.