LIVEDESCRIBE – CAN AMATURES CREATE QUALITY VIDEO DESCRIPTION

by

Carmen Branje, BASc (Ryerson University, 2006)

A thesis

presented to Ryerson University

in partial fulfillment of the

requirements for the degree of

Master of Management Science

in the Program of

Management of Technology and Innovation

Toronto, Ontario, Canada

I hereby declare that I am the sole author of this thesis.
I authorize Ryerson University to lend this thesis to other institutions or individuals for the purpose of scholarly research.

Carmen Branje — 01/08/08 Date

I further authorize Ryerson University to reproduce this thesis by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

Carmen Branje — 01/08/08 Date

LiveDescribe –

Can Amateurs Create Quality Video Description

Master of Management Science, 2006

Carmen Branje

Ryerson University

Abstract

This thesis explores amateur video description facilitated through the video description software program called LiveDescribe. Twelve amateur describers created video description which was reviewed by 76 sighted, low vision, and blind reviewers. It was found that describers were able to not only produce description but that their descriptions seem to be perceived as having an acceptable level of quality. Three describers were found to be rated as "good", three were rated as "weak" and the remaining six were in a "medium" category. The common factors that appeared to characterize the good describers were a soft non-obtrusive voice, a moderate amount of well placed descriptions, moderate description lengths and English as a first language spoken without an accent or regional dialect. It was found that LiveDescribe was a useful and easy to use tool and that it facilitated a video description work flow process for amateur describers.

Acknowledgements

I would like to thank many people, without whom I would not have been able to complete this research and my degree.

First and foremost I would like to thank my thesis supervisor Dr. Deborah Fels who saw potential in me as a young undergraduate student and sparked my interest in science and academic research. Over the last several years she has been my guide and my mentor and I could not have asked for a more dedicated, intelligent and selfless individual to lead my academic pursuits.

I would like to thank my parents Maria and Tony Branje and my partner Vanessa Shepstone. Without their constant emotional, moral and financial support I would not be where I am today.

A special thank you to the wonderful amateur describers who gave up part of their day to create video description and to the hundreds of sighted, blind and low vision participants who took the time to complete my video description survey. Your efforts have not only helped my research enormously, but will in no doubt help the blind community as well.

Thank you to all my friends and colleagues who provided a wonderfully stimulating intellectual environment to work in and around, especially J.P. Udo, Bertha Konstantinidis, Maria Karam, Danny Lee and Emily Price.

Finally, I wish to acknowledge the funding support provided by Canadian Heritage and NSERC.

Table of Contents

List of Tables

List of Figures

1. Introduction

Media and culture are important aspects of the growth and development, and communication structures of a society. It is important that all members of that society have access to the media and cultural artifacts so that they can be equal participants in the advancement of it. However, some members within society such as those with sensory disabilities (e.g., low vision, blindness, hard-of hearing or deafness), can be excluded because cultural materials are often produced for multimedia displays and require vision and hearing in order to access and consume them.

There has been some progress in creating more accessible culture through closed captioning and video description, and various attempts to produce inclusive original works. However these attempts are only a good beginning, but remain incomplete solutions; much more work is still required.

Video Description (also referred to as DVI, described video or audio description) is a new technology that has been developed to provide access to television, film and theatre content for viewers with vision impairments. Video description "provides spoken description of the visual content such as costumes, facial expressions, and sources of sound effects" (Fels, Udo, Diamond, & Diamond, 2004, pg 1) through a second audio channel. The description is created by a trained, professional describer through a conventional process o f script writing, recording and audio mixing and then inserting it into the spaces between the dialogue elements of the content. The timing and precision of the description is critical as the available spaces are often short in duration (less than 5 seconds) and infrequent. As a result, not all of the important or relevant visual information can be described. Description then becomes an interpreted process where

decisions regarding what visual information is important to include or exclude and the words to describe that information must be made.

Like captioning and other accessibility technology, video description in many countries is regulated by a broadcaster regulator. In 2004, the Canadian Radio-Television and Telecommunication Commission (CRTC) mandated, as part of broadcast licenses, two hours of video description per week until 2008 and then it is increased to four hours per week for most broadcasters. (CRTC, 2004). While the CRTC should be applauded for taking these first steps toward increased accessibility to television, the requirements are modest at best. It seems to suggest that the CRTC is attempting to introduce video description slowly perhaps to avoid imposing too high a financial burden on the industry or to wait until further costly infrastructure is in place, suggesting that the "bottom line" takes a priority over access. Couple this with the fact that the CRTC rarely, if ever, revokes broadcasting licenses due to non-compliance with accessibility mandates and blind audiences are left with content that for the most part has little, if any, description.

Video description is a young discipline, and there is very little research supporting description strategies and best practice guidelines. In addition, the television, film and live theater industries are scrambling to meet regulatory and legislative mandates for video description with a limited knowledge and evidence base, and few technologies and tools and infrastructure support systems.

There are also a number of issues that impede the development of video description techniques and technologies. The Independent Television Commission, 2000 (ITC) reports that "on average it takes one describer a working week to produce between one and a half and two hours of described programming" (The Independent Television

Commission, 2000, pg 12). This then translates into increased cost and production time for creating and deploying a piece of content. This issue is particularly important for programming that is broadcast daily such as *The Daily Show* (Fishman, 2006) or soap operas, where a there are tremendous time constraints placed on the production process – a one-week lead time to add video description is often not feasible. As a result, the industry is reluctant to embrace a new process that only increases the production costs and adds little perceived value other than to meet a government mandate.

A second important issue for video description is that because of the cost and time required to create it, and the limited regulatory motivation, there is relatively little content described. People who are blind do not want to be limited in their consumption of content to only what broadcasters decide should be described but want access to as much content as possible.

One method of increasing the efficiency of the description process is to introduce more automatic processes such as automatically detecting the non-dialogue elements, or automatically detecting, characterizing and describing salient visual information to reduce the time it takes to create a video description track. A tool that has recently been developed to implement some of these automatic processes, and that is the focus of this thesis, is LiveDescribe (Branje, Marshall, Tyndall, & Fels, 2006). It is a software tool, developed at the Center for Learning Technologies at Ryerson University to support different forms of description including post production, "near-live" description and full live description. LiveDescribe allows a user to import video content, add additional audio narration and then export a completed video containing overlaid description. Three different signal processing algorithms are integrated to detect periods of non-dialogue

(insertion point and length) where description can be inserted. A zero-crossing based measurement, entropy, and root mean square (RMS) are used to detect the non-dialogue spaces. Spaces not detected by the automatic system can be manually defined by the describer (see section 3 for a complete system description).This software thus has the potential to reduce the overall time required to produce description because of the automated processes it uses.

Another important element of video description that has generally been overlooked in the production and research communities is the fact that blind viewers have been attending cultural events before video description services were ever formalized. The majority of video described events has been and still is being carried out on an ad hoc and personal basis by family members, friends, and associates of the blind or low vision person. Is there a way to support this method of description so that people can capture and share these personal versions of descriptions thus increasing the corpus of described content?

The recently emerged wiki phenomenon made possible by the internet and spear headed by the collaborative online encyclopedia Wikipedia, has opened the door for new ways of computer supported, human to human collaboration. Individuals across the globe are creating and sharing vast quantities of multimedia content through wiki technologies. Individuals can consume, respond to, edit or concatenate new content so that the content becomes an organic system that evolves by public debate and contributions. Although there may be some question of quality control for this content, the collaborative and public nature of it ensures that it can be challenged, upgraded or debunked.

Wikipedia has become an important source of knowledge and information that is easy to find, and is freely and publicly available. This concept has been embraced in the Web 2.0 community and as a result, wiki websites are becoming commonplace for all different types of applications (e.g., group collaboration in research projects).

Using tools such as LiveDescribe and Web 2.0 technology, capturing, preserving and making public this type of ad hoc description may be possible. I propose that a public video description wiki that provides people such as the friends and family of people who are blind or low vision or others with an interest in producing amateur video description, a venue to create, store and make available the descriptions they create. Descriptions for one particular show then become entities that can be changed, updated, and edited so that they improve in quality and quantity, or people can also introduce additional descriptions (called extended descriptions) to supplement existing ones.

This thesis presents the results of a study conducted to determine the feasibility of amateur video description creation and of the usability of LiveDescribe to facilitate it. This study is the first step in determining the possibility of a video description wiki. The focus of this research will be to explore the viability of non-professional users creating description with an acceptable quality level. Sites such as Wikipedia and YouTube provide evidence that non-professional users can create and manage "acceptable content". Further evidence that suggests this task is potentially feasible is the fact that people have been describing visual elements of movies, television and theatre to their blind relatives and friends. In addition, the use of LiveDescribe to facilitate the process of creating description by amateurs is considered in this thesis.

This thesis is as an exploration of the various aspects involved in amateur video description creation and the usability of LiveDescribe. It will not directly address description style, captioning, personal description style preferences, description delivery methods, opinion differences between description of different episodes, different shows or different genres of programming, although the issues may appear during the exploration process.

In this thesis, section 2 provides a review of the literature on video description, and web 2.0 technologies. Section 3 outlines the LiveDescribe description system. Section 4 will outline the research questions of this thesis and the method used to investigate the answers to those questions. I will report all findings and data analysis in section 5. The results reported in section 5 will be discussed in section 6. Section 7 will discuss the limitations of this research and finally, section 8 will conclude the thesis and establish future research suggested by the findings of this thesis.

2. Literature Review

2.1 Video Description

The production of video description is currently a process that occurs after the entire show has been produced and is ready for distribution or broadcast. A third party description company such as Audio Vision Canada usually receives a copy of the show and is asked to produce description without the input of the creative team responsible for the product. In this process, the description provided is in the third person narrative form and is usually delivered without flare or adjustment for the tone or mood of the program being described. Descriptions are created using a process that is mostly manual, involving the description writer watching the entire program once through, sometimes with a first pass without video to understand the show. Spaces between the dialogue elements of the show, where descriptions can be inserted are then manually identified. A description script is then written to fit into the identified spaces. A describer (who is often the same person as the description writer) is brought into a professional studio to record the descriptions which are then edited into a secondary audio track. Adjustments for volume levels, speaking rate and timing are made during this editing process.

Currently there exist several software based video description tools to assist in the creation of video description, including Magpie (NCAM, 2003), Softel (Softel-USA, 2001) and Capscribe (Treviranus, 2004). These tools will not be discussed in this thesis as a discussion regarding them does not fall within the scope of the research questions posed. Future study is needed to compare the usefulness and ease of use of different video description tools.

2.1.1 Need for description

The latest data from the 2001 Statistics Canada publication The Participation and Activity Limitation Survey (PALS) showed that 3.1% of Canadian females and 2.0% of Canadian males identified themselves as having a "seeing disability" defined as "difficulty seeing ordinary newsprint or clearly seeing the face of someone from 4 meters (12 feet)." (Statistics Canada, 2001) With an aging baby-boomer population this number will likely increase over the coming decades. The demand for services such as video description, designed to help those with visual disabilities access television, film and live production content, will only increase with these changing demographics.

Schmeidler (2001) found that people with visual impairments enjoy television programming just as often as those who are sighted. "People who are visually impaired watch television at about the same rate as does the general public."(Schmeidler & Kirchner, 2001, pg 1). However, just because people who are blind or low vision watch television does not mean that they are using video description or that they are understanding or seeing all of the show and experiencing the same entertainment value as those with full sight. The fact that vision impaired people enjoy television and seem to want to consume it as much as others provides motivation for the provision of video description services.

Along with an increased video description production demand fueled by an aging and growing population is a demand and need for guidelines and processes that assist describers in producing descriptions effectively and efficiently. It is reported that production costs for video description are approximately $4,000 per hour of content (Clark, 2001), meaning a two hour movie could be as high as $10,000. Since content

producers are under constant financial pressure, the cost of video description represents a large barrier in the creation of a high volume of described content.

2.1.2 Description Does Add Value

Peli and Fine (1996) showed that low vision and blind users are indeed missing a significant amount of information when viewing television programs. It their study participants were showed short ten minute segments of two television shows, one mystery program and one nature documentary program to 25 low vision participants, 24 sighted participants and third group which heard the audio only portions (to simulate blindness) of the show. A multiple choice questionnaire was given to participants after they had viewed each program to determine the amount of visual information delivered. The study showed that those that heard the audio only programs answered fewer questions correctly than did either the sighted or low vision groups. This suggests that audio description may be suited to fill in the gap left by an absence or reduction of video information and that providing description of the visual details of documentary content could be beneficial to blind and low vision viewers. This study was only meant to explore how much information low vision audience members understood from non-described content. No actual description was included in the study.

Schmeidler and Kirchner (2001) conclude that video description does indeed add value to programs for people who are blind or low vision. By surveying the viewing experience of over 100 blind or low vision participants, Schmeidler and Krichner found that "individuals recount that dramatic difference they experienced when they saw a described program or performance for the first time." (Schmeidler & Kirchner, 2001, pg

209) Low vision and blind participants were shown two documentary science programs with relatively contrasting styles. The first, *Orphans of Time* from the *New Explorers* series and the second *Wild Dogs of Africa* from the *Nature* series. *Nature* had a single narrator without any other human voice and contained longer segments of visual-only content while *Orphans of Time* had much more dialog, featured many narrators and participants speaking amongst themselves.

Participants viewed the entire first documentary lasting about 40 minutes and then answered questions regarding the "facts" of the program. They then watched half of documentary 2, lasting 40 minutes - the same length of time as documentary 1, and were then again asked questions about the facts of the show. Half of the participants viewed program 1 with description and program 2 without while the other half viewed the reverse. Telephone interviews were conducted prior to the study in order to collect demographic information such as vision status and television viewing habits. The group that received video description did significantly better on their questionnaire than the group that received no description. The researchers found that viewers were not only more informed about the content of the video described program, but they took away a greater emotional experience and were better able to share their experience with sighted friends or family. However, in this study only two documentaries, lasting 40 minutes each, with professional description were evaluated. A novelty affect with the video description, the style of content and the relatively short duration of exposure may have resulted in overly positive outcomes. In addition, documentary content is fact-based and may lend itself better to certain styles or types of video description. Any conclusions

drawn may thus only be relevant to this type of content in this particular genre of programming.

Pettitt, Sharpe and Cooper (1996), in their study of AUDETEL (Audio Described TELevision), discovered that "AUDETEL goes a long way to improving access to television by enabling visually impaired people to 'watch' programmes which would otherwise be inaccessible and by enhancing those which are already watched."(Pettitt, Sharpe, & Cooper, 1996, page 6). This study offers insights into the value of description over multiple genres of programming as well as benefits of description as a long term service. One hundred low vision or blind participants took part in a field study where they had an opportunity to watch 4-5 hours of video described programming over a five month period of time. Quantitative data specific to the television viewing habits of the participants as well as qualitative data such as participant comments were gathered. The five month time period of the study and relatively large sample size of 120 blind and partially sighted people, showed that description was not a novelty and did not lose its value over long periods of use. The variation of program genres studied in this research also showed that there was value in description across most types of programming.

Benefits of description should not necessarily be limited to pre-recorded video and for people who are blind. Peters and Bell (2006) suggested that video description can be used to enhance the experiences of many when viewing digital photographs. Pfanstiehl and Pfanstiehl (1985) showed the value of description in live theatre productions. "Audio description has proved to be well worth all the effort involved, as the numerous appreciations received from visually impaired theatre-goers have been

shown."(Pfanstiehl & Pfansteihl, 1985, page 17). Although these are anecdotal reports, they do at least suggest further study in the area is warranted.

Schmeidler (2001), noted that informal amateur description delivered by a friend or relative has been shown to add value to a video or presentation. "Informal description by others who are watching the programs clearly enhanced television viewing for almost all the people who experienced it."(Schmeidler & Kirchner, 2001, pg 210 Schmeidler (2001) and Fels (Fels, et al., 2006) found that people who are blind or low vision are more comfortable talking with sighted people about the show (the content, their impressions of it, their likes and dislikes of it) when it has been described than when it has not. This is likely because video description allows for a more profound understanding of the program and an increased level of comfort with the content. Research at the Center for Learning Technologies at Ryerson University is being carried out to explore questions regarding description style and form. This research has been studying the differences between the entertainment and information value of description for different genres, users and styles. Fels et al. (2006) have reported some preliminary research into the possibilities of different styles of description; specifically the impact of first person description rather than the traditional third person omniscient style. It was found that the first person description was less trustworthy but more engaging. It was also suggested that the show described with a first person description was "fun to watch and made viewer's experience with the show interesting." (Fels, et. al., 2006, pg 81). Although this study provided some new evidence regarding the importance of entertainment to viewers, there were also some important limitations. There were only 7 participants employed in the study and more than half of these people were not within the

target age demographic for the show. As a result, some of the comments and opinions expressed by the participants may have been influenced by their attitude toward the show rather than the video description. In addition, as with other studies, only short clips of one genre of show, an animated comedy, were used.

To address the issue of stylistic preferences and entertainment, Fels et al. (2006) showed that having the creative team produce the video description content during the course of an original production was an alternative approach that was preferred over the conventional process. Seven blind and seven sighted individuals watched four clips of the cartoon Odd Job Jack, each two to three minutes in length. Each clip was presented using a different style of description, either no description, conventional third-person or a new first-person narrative style. The study demonstrated that not only did blind audiences prefer the first-person style but that sighted audiences also enjoyed watching the show with this style of video description. They would even consider purchasing the audio-only track of a show with this type of video description for listening to while in the car or commuting. While this study shows some interesting preliminary results, there were several important limitations. Only one show, Odd Job Jack, from one genre, adult cartoon, was studied and participants only experienced the content for a short period of time.

In addition, in all of the studies presented in this literature, the describers were professionally trained and produced carefully controlled and high quality video descriptions using only one style. There was little to address the notion that video description can be provided by amateur describers who are the friends and family of blind and low vision viewers even though there is an acknowledgment that this type of

description is prevalent and useful to blind viewers. It may be possible to use new Web 2.0 technologies to enable amateur video describers, and make description more widely available. However, there is no evidence to support or deny this assertion.

2.2 Web 2.0 and Wikis

Web 2.0 has recently been introduced as the latest web innovation that is set to revolutionize internet technology. Web 2.0 is a concept of the World Wide Web where users create, edit and share multimedia online content (Bleicher, 2006). One early and prominent implementation of this concept is the wiki, exemplified by the popular wiki encyclopedia Wikipedia.com (Long, 2006).

A wiki is a publicly accessible reference site that can not only be viewed by anyone but can also be edited by anyone. It is based on anti-authoritarian concept as it gives no credit for credentials or education. Editing is carried out through a community revision process rather than a single point of authority. For example, any person can edit existing entries or create new ones in Wikipedia. The result has been one of the Internet's busiest sites containing over 2 million articles on a very wide range of subjects. One advantage of this concept is the ability to amass large quantities of publicly generated content and make it widely available with little production cost to the hub or central repository which distributes the final product. This advantage has the potential to carry over to a video description wiki and result in large quantities of relatively high quality description, created by volunteer non-professional describers. However, use of wiki technology for video description should be informed by research carried out on web 2.0

technologies because there are numerous issues such as quality control, data trustworthiness, security and copyright issues that must be considered.

Stvillia (2005), through the study of Wikipedia discussion and change history pages, described the quality control procedures in Wikipedia as being very strict and thorough. Any member of the public could contribute Wikipedia content but that content was scrutinized, reviewed and discussed by fellow community members before it was accepted as content. (Stvilia, Twidale, Gasser, & Smith, 2005).

This finding suggests that although there is no formal overarching authority responsible for quality control, there is an emergent quality control mechanism realized by the community of content providers and viewers and facilitated by the site itself, and that this mechanism can be relied upon to provide an acceptable quality level for the content provided. It is hoped that by simply providing a mechanism for the video description community to filter or rate uploaded descriptions, that little, if any quality control efforts have to be taken in order to ensure high quality description is accessible on the site. High quality description which is rated high by the community will filter to the top, while poorly rated description will sink to the bottom.

Ramos and Piper (2006) suggest that librarians and have little to lose by embracing these new information creation and sharing tools. Furthermore, they found it possible that wikis "can yield information found nowhere else." (Ramos & Piper, 2006, page 3) This could be due to the fact that the information is created for free and that information that would not normally have a business incentive for existence suddenly receives the resources required to bring it about. This finding is of particular interest since it means that wikis have a tendency to produce content not found elsewhere, and

therefore it is hypothesized that a description wiki will produce description not found elsewhere.

The results from these studies seem to indicate that amateur content creators can make valuable, accurate and unique contributions to the public knowledge base. It appears that quality control of content can be adequately managed and maintained in the public domain by the public investors in that content.

Web 2.0 technologies, particularly wikis, have inspired the research that I propose to carry out for my thesis as they have shown that quality control is manageable by the public in a public domain. In addition, video blogs and public video spaces such as YouTube provide evidence that people want to produce and consume amateur video materials.

Providing user created access seems like a feasible and valuable undertaking, however, as is often the case with immerging information technology, copyright law can affect the free development of such a system. A description wiki would only store user generated descriptions and not the original program to be described, and may therefore fit within current copyright laws. The combining of the description and the original audio track may breach copyright law but that legal issue is outside of the scope of this thesis. Legal research would be required to investigate the ramifications of combining the original audio track and additional description tracks.

2.3 Theoretical Foundations of the Approach

Due to the limited research on video description and the fact that it is such a new process, no formal conceptual frameworks for description exist. However, the tenets of universal design theory can inform the practice of video description and the design of LiveDescribe. Universal Design Theory is a set of principles set out to create a more a useable and accessible environment for everyone. The seven principles of design are (Story, 1998):

- **Equitable use**
  Make designs appealing to all users.
- **Flexible use**
  Provide choice
  Facilitate the user
- **Simple and intuitive use**
  Keep it simple
  Be consistent
- **Perceptible information**
  Deliver information in different modes (pictorial, tactile, verbal)
- **Tolerance for error**
  Fail safe
- **Low physical effort**
  Neutral body position
  Minimize repetition
- **Size and space for approach and use**
  Accommodate variations in size
  Accommodate for assistive devices

Although many of these design principles are specifically targeted at physical construction design, most can also be applied to the design and construction of LiveDescribe and to the practice of video description itself. This framework will be used to guide for the completion and further modification of LiveDescribe in the following ways:

1. **Equitable use**

   Although this is a challenging area since vision and hearing are required to create description, issues of accessibility should not be ignored. Zooming features have been integrated into the timeline of LiveDescribe to accommodate low vision describers. Other features such as keyboard access to help those with limited physical mobility will be implemented in LiveDescribe as development progresses.

2. **Flexible use**

   Flexible use can already be seen in LiveDescribe by its two pronged approached to creating description. Describers can chose to use the timing or automation features of LiveDescribe or they can chose to ignore them and take a more manual approach to the process. Likely expert describers will use less of the automatic tool, while novices will tend to use them more. In addition, LiveDescribe can use virtually any audio source or pointing device.

3. **Simple and intuitive use**

   This principle has been at the forefront of the design of LiveDescribe. The timeline interface was chosen for its familiarity to multimedia artists. In addition the main functionality of LiveDescribe has been kept to one task, description.

4. **Perceptible information**

   The main feature of LiveDescribe is its ability to collect and relay data back to the describer so this principle of design has been followed carefully. From the very important timeline, to the space list, and information display window LiveDescribe is constantly delivering information in different modes.

5. **Tolerance for error**

   LiveDescribe is designed with the use of exception handling. Any errors thrown by the system will be caught and handled gracefully, alerting the user to the problem, allowing the user to save work, the close the program and continue at a later time.

6. **Low physical effort**

   Although a customized physical input device is not yet available for LiveDescribe, it is in the future version plans for the software. Keyboard access is something than can be implemented during the development phase of customized input devices.

7. **Size and space for approach and use**

   LiveDescribe is displayable in many different resolutions, and the functional windows can be resized and moved.

2.4 Description Standards and Guidelines

Existing guidelines and standards specific to video description that do exist seem to be based on anecdotal evidence rather than empirical research. In reality what exists is a de facto set of standards developed over the course of video description's twenty year history by various bodies producing or legislating video description.

For example the Independent Television Commission (ITC), which no longer exists, and was superseded by the Office of Communications (Ofcom), was a UK commercial television regulator which released a document called Guidance on Standards for Audio Description (The Independent Television Commission, 2000). It states that "there are three golden rules to description: describe what is there, do not give a personal version of what is there and never talk over dialog."(The Independent Television Commission, 2000, pg 9)

In the ITC document, there are various standards and guidelines for producing video description which include technical issues such as choosing shows, preparing scripts, and program and recording audio volume, as well non-technical issues such as which tense a describer should use, the use of proper names and pronouns and prioritizing information. In addition, the ITC guide gives recommendations specific to various genres of programming, such as musicals, soap operas, sports and live events and documentaries. The guide describes the skill sets of a good describer needing "good writing skills, a clear, pleasant and expressive voice and a thorough knowledge of the needs of a visually impaired audience."(The Independent Television Commission, 2000, pg 8)

Although these standards seem reasonable, they have yet to be been tested by rigorous academic research. Fels et. al. (2006) showed that the personal description of the describer should be incorporated into description it is actually useful and desirable (Fels, Udo 2006) which seems to be contradictory to what the standards claim is acceptable.

From this literature analysis comes several important discoveries that suggest further study is warranted. First, although the research on description is limited, it has been shown that video description is a worthwhile endeavor as users of video description appreciate and benefit from its existence and use. Secondly, it has been shown that amateur description, at least in an informal setting, is plausible. Thirdly, it is clear that wikis are a very cheap and efficient source of massive amounts of original, relatively high quality content. Lastly, the inability of government regulatory bodies to enforce mandates for video description creation suggests that alternative sources of video description must be sought. All of these discoveries combined suggest strongly that research into a video description wiki, facilitated by LiveDescribe is a worthwhile undertaking.

2.5 Gap Analysis

As research in video description has only appeared over the last 10-15 years, it is still a new discipline. As a result, there exist many gaps in the body of description knowledge. Many aspects of descriptions, such as style of description, the type of language which should be used, the impact on users in the long term, and how descriptions fits with different genres remain to be studied. However, the industry is demanding solutions and systems today and thus are proceeding standards and guidelines

with limited guidance from research. Unlike closed captioning, which is a verbatim equivalent of spoken dialog, video description is open to interpretation and unique styles. Different describers may produce different descriptions for identical content. Users may also have distinct preferences regarding the style of description they like and want to hear, and this may also vary with genre. One of the fundamental principles of universal design states the user preferences should be considered and accommodated. Can video description offer this level of viewer support? Is it desirable, effective and efficient to provide?

Questions arise regarding the different styles and forms that are possible and whether viewers have different preferences for those as well. Should descriptions be delivered in a manner of fact, "news reader" style, or should the emotion and spirit of the program be reflected in the tone of description? What narrative form should be used, should a character present in the show present the descriptions, or should it be a narrator outside of the story universe. "At present, there is neither a voice of the consumer nor a body of empirical research to indicate which approach may be better." (Piety, 2004, pg 467)

One of the major gaps in video description research that I have identified is that only professional, experienced and well-trained describers have provided the descriptions for evaluation. However, I suggest that non-professional/amateur describers can also provide video description because of the evidence provided by wikis and other Web 2.0 technologies. Traditionally, the task of encyclopedia authoring was left to a small group of professional, authoritarian, elites. Wikis such as Wikipedia however, are breaking this long standing practice. If masses of amateur researchers and writers can create an

encyclopedia full of useful information on a site like Wikipedia, it is possible to suggest that amateur describers could create video description content that is of sufficient quality and can be understood and useful to viewers who are blind or vision impaired.

Although Schmeidler (2001) found that informal amateur description can provide value to a program and may help to overcome the lack of description available she also suggested "having to describe—or needing someone to describe—may be an unwelcome distraction or burden"(Schmeidler & Kirchner, 2001, pg 210). Schmeidler did not explore the concept of the amateur describer in any detail and this area of description warrants more study.

Another gap in current knowledge is whether a web 2.0 collaborative structure such as a wiki can be designed to support access by creators and users of video description, and whether that wiki can support the universal design concept of options and preferences.

Unlike traditional television video descriptions, which is accessible essentially through a single button press of a television remote, blind users may have more difficulty accessing description provided through a wiki/website. One solution would be for blind or low vision users to have sighted friends or family access description for them, however, low vision or blind people may not want to depend on this for access to description. Another possible solution is through screen reader tools designed to give internet access to the vision impaired. Screen readers, such as Freedom Scientific's JAWS™, GW Micro's Window-Eyes and Alva Access Groups OutSpoken, are designed to provide access to computer applications such as browsers and office productivity software by capturing text as it is written on the computer screen and then speaking it to

the user (Sierkowski, 2002). Once an online environment, based on the findings of the research outlined in this thesis, is constructed to support descriptions, evaluations with screen readers and blind users is a potential next step. However, it is outside of the scope of my research proposal.

2.5 The Need for This Project

The current catalog of video described material is very limited because the high cost of production relative to other services such as closed captioning. (Peli, 1996) Currently, only the most popular movies and television shows are video described, meaning only a very selective catalog of films are available to those with visual impairments. For example in the action category, there are only about 60 titles available from AudioVision.com (the largest retailer and producer of video described content). In addition, there are very few outlets for renting or purchasing video described content, making it difficult for those with visual impairments to obtain what little described content is out there. Finally, due to the relatively high cost of producing video description, and the relatively small audience/market, which benefit from video description, there are few, if any, business incentives to justify creating more video description than legislation mandates. Video described DVD's purchased from the AudioVision.com website cost $29.95 US plus shipping charges, a cost that is higher than many commercial DVD's purchased from traditional retail outlets. There are also few regulations in place throughout the world to require producers to provide description. I then suggest that amateur description that can be made publicly available through

services such as wikis and supported by easy to use software tools is one possible solution and the direction of my research.

3. LiveDescribe Description

LiveDescribe is software designed to facilitate the creation of video description for digital video content. The interface is designed to support professional and amateur describers with the ability to input text for descriptions, record, insert, edit, and manipulate descriptions on a timeline, and view graphical representation of the audio track including the spaces available for inserting description. The non-dialog spaces available are automatically detected and graphically represented on a timeline. A detailed description of LiveDescribe is included in this chapter along with the conceptual approach to the coding process and user interface that was developed.

3.1 LiveDescribe User Interface

The LiveDescribe user interface is modeled after many currently existing media software packages, such as Adobe's Premier or Apple's Final Cut Pro which use a track or timeline-based conceptual model. This type of interface creates a description development environment that has features which are familiar to many users who work with video or audio data. Taking advantage of that familiarity may reduce the learning required to successfully manipulate the interface and increase the ease of use of LiveDescribe, particularly by amateur describers. In addition, since video has a time dimension, representing the data in this way fits conceptually and functionally.

Most of the important features of LiveDescribe, such as video navigation, space definition and description recording are accessible through the main timeline, or in

conjunction with another window. A more detailed description of each interface element and the features of LiveDescribe follow.

As illustrated in Figure 1, LiveDescribe is comprised of five main windows from which the majority of the functionality can be accessed. The windows are as follows:

1. Video Window
2. Timeline Window and Timeline Legend Window
3. Video Navigation Window
4. Video Information Window
5. Description and Spaces Window

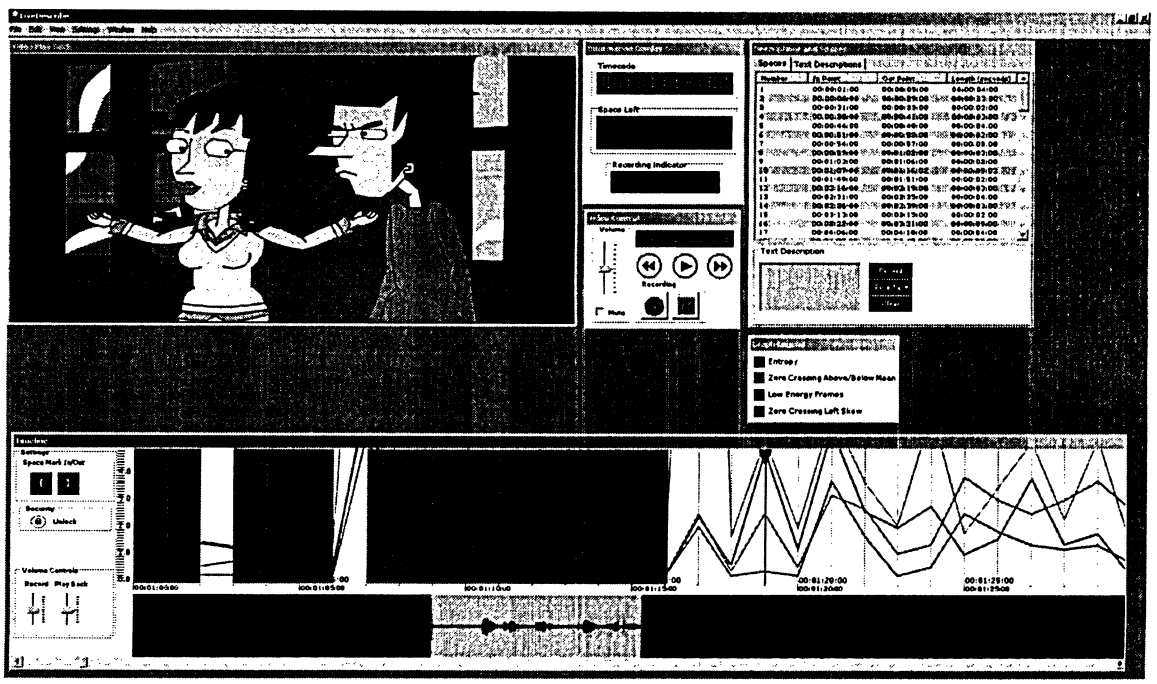


Figure 1 - Screenshot of main screen of LiveDescribe software. Shown on the far left are two unselected spaces, while the rightmost space is selected. Also, the selected space is highlighted in the Description and Space List window in the upper right corner of the screen.

1. **Video Window**

The main purpose of this window is to view video playback. It can be resized and repositioned as needed. The video to be described can be loaded into this window and controlled using the video navigation window.

2. Timeline Window

The timeline represents the most crucial aspect of LiveDescribe's user interface. From the timeline, the video can be controlled by selecting and dragging the triangular-shaped position indicator. As the video is played, fast forwarded or reversed this position indicator moves to reflect the current position in the video.

Description spaces (seen on the timeline as green, and the selected red square) are automatically determined by the discrimination algorithms. The purpose of these spaces is to show the describer potential description opportunities and the duration of the available space for inserting those descriptions. Since the speech/music discrimination algorithms are not completely accurate, users may want to adjust spaces or insert new ones. Spaces can also be edited, fine-tuned and manually defined by the user within the timeline.

The timeline also serves as an important information portal. The lines seen throughout the timeline represent the results from the dialog discrimination algorithms. Users see a visual display of the dialog/non-dialog analysis and can understand the flow of the audio in the content through this display. A valley created by all the algorithm lines usually represents a period of non-dialog, while all others are most often dialog.

As seen in Figure 4, directly below the timeline is the description track. It appears in the figure as black line with a blue wave form directly below the red highlighted

section. When a description is recorded it will appear as this blue wave form under the selected graph section. The describer can reposition or delete recorded descriptions by clicking and dragging the description across the lower area of the description track.

The buttons on the left panel of the timeline window are the space mark in/mark out buttons for defining a new description spaces. When a user wishes to define a new space, he will move the cursor to the desired position, click mark in, then move to the desired out position and click mark out. In addition, description playback and recording volume can be set using the sliders located in this panel.

The Timeline Legend window contains the graph legend which defines each of the graph lines shown in the timeline.

3. Video Control Window

The video control window also allows a user to navigate through the video displayed in the video window. Users can control volume, play, pause, reverse or fast forward as well as start and stop the recording of description from this window. This window has standard functionality and interface objects that are similar to many standard video or audio players or recorders.

4. Information Display Window

The information display Window is an information panel only. It also displays the current position in SMPTE timecode of the video as well as a recording indicator which flashes on when recording is in progress. In addition, it contains a "Space Left" bar which serves to indicate to the describer how much of a description space remains once they begin describing. Once users begin describing from within a space, the Space Left bar shows the time remaining in that space using a progress bar.

5. Description and Spaces List Window

The description and spaces list window, has three main functions: a search and navigation function; a writing utility, and a recording device. The search and navigation function provides a numbered list of all the automatically and manually defined description spaces available throughout the video content. By clicking on one of the space listings, the video will automatically queue to the beginning of that space. This allows users to quickly browse through the video content and review automatically or manually defined description spaces so that they can find parts of the video which most likely require description. This feature also allows the user a choice between a timeline based workflow (in the Timeline window) and a timecode or listed based method.

In the description and space window, the user may also type in a text description and associate it to a description space through the text box provided in this window. This provides not only a location for the describer to write down descriptions, but it also provides a valuable association of that written description to a portion of the video which will eventually contain the recorded version of that text. Typing in the description also reports the required word per minute rate necessary to fit that description within that space, to the describer, so that an estimation of how fast a description must be delivered can be obtained before recording. Having this timing information can help the describer be efficient in planning their descriptions because they can avoid the manual approach to description timing.

Clicking the record button brings up another window with the written text in a large readable font. When the user begins his recording by clicking record, the text will be highlighted one character at a time (similar to that seen with a karaoke singing

machine). The timing of this display is such that it will fit into the available space designated for that description. This highlighting feature functions as an indicator to the describer as to how fast or slow he should read the prewritten description in order to fit that description within the designated space.

3.2 Development Platform

LiveDescribe was developed using Microsoft Visual Studio, specifically the programming language C#. I chose this development platform because of its rapid prototyping functionality and its ability to directly interface with Microsoft DirectX and more specifically DirectShow. DirectShow allows the reading, writing and manipulation of most types of digital media, and in almost any compression format. DirectShow uses a centralized repository of video and audio compressors and decompressors (CODECS), meaning that common high level source code can be used across all types of media, greatly reducing development time and complexity. DirectShow also uses a filter based conceptual framework allowing for easy manipulation of media data. Media data processing is conceptualized into filters where each filter is responsible for a specific task such as parsing a movie file into video and audio streams or parsing and decompressing a compressed audio stream into a readable uncompressed format. The data from the filter is then passed to the next filter which performs the next specific task.

An example of a DirectShow filter graph can be seen in Figure 1 which separates MP3 audio data from a AVI video file (.avi) and converts it to a WAVE format (.wav). Data (thedailyshow.avi) moves from the file source filter through to the AVI splitter filter which splits the stream it receives into separate audio and video streams. Once separated,

the data then moves to the next filter, which it then passes down the graph to the next filter (MPEG layer 3 Decoder). From there, the audio signal is sent to an mp3 compression filter on so on, until the data packets reach the file writer filter at the end of the graph. The Sample Grabber filter is a customized filter which contains the dialog discrimination algorithms. This conceptual framework allows for plugging and unplugging of filters to support the manipulation of the video or audio signal at any point along the filter graph.

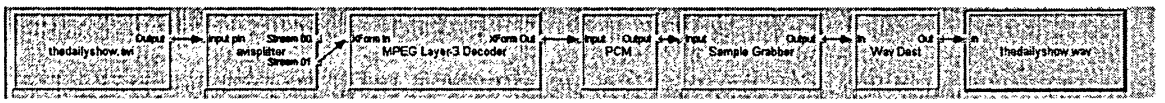


Figure 2 - A DirectShow filter graph for seperating wav audio from an avi video

3.3 Discrimination Algorithms

LiveDescribe uses three relatively simple algorithms to discriminate speech and non-speech audio. I selected these algorithms because they provide the basic functionality and redundancy necessary for the discrimination process and their implementation is relatively straight-forward. The three algorithms used include an entropy measurement algorithm, zero crossing and one based on the root mean square (RMS). When combined the algorithms make for a relatively robust speech/non-speech discrimination system due to redundancy.

Speech is created by joining several bursts of sound or words, into a string of bursts we call a sentence or phrase. "The characteristic structure of speech is a succession of syllables composed of short periods of frication followed by longer periods of vowels or highly voiced speech. This simple but widely accepted generalization can be used to

positively recognize a waveform as being predominately speech."(Saunders, 1996, pg 994) As seen in Figure 2, which is an example of a speech waveform, these bursts of sound appear and dissipate quickly and are followed by equally brief and short lived periods of silence. This rapid succession of interwoven bursts and silences creates a "peak and valley" graphical waveform which can then be detected using algorithms such as zero-crossing or entropy measurement (Pinquier, 2002), which are employed in LiveDescribe.

As seen in Figure 3, music is often very uniform in the shape of its waveform (Pinquier, 2002). Professional audio engineers work hard to mix, adjust and compress audio signals and create an even waveform that is pleasant to the ears of a listener. This uniformity in shape can also be easily detected by the entropy and zero crossing algorithms used in LiveDescribe.

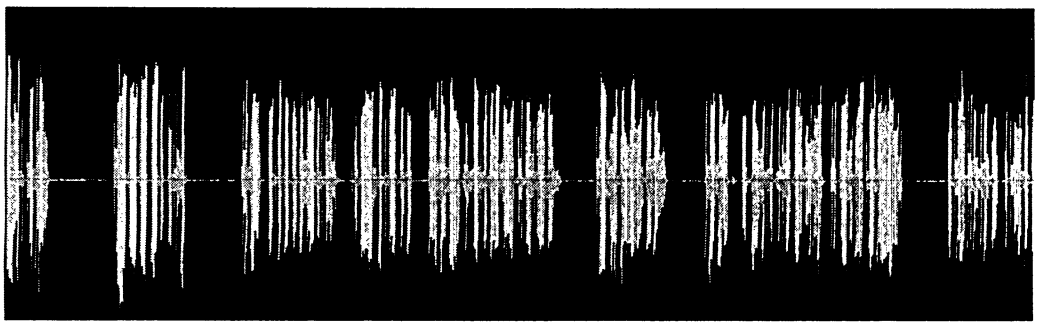


Figure 3 - An example of speech waveform. The first minute from a reading of Charles Darwin's Origin of Species

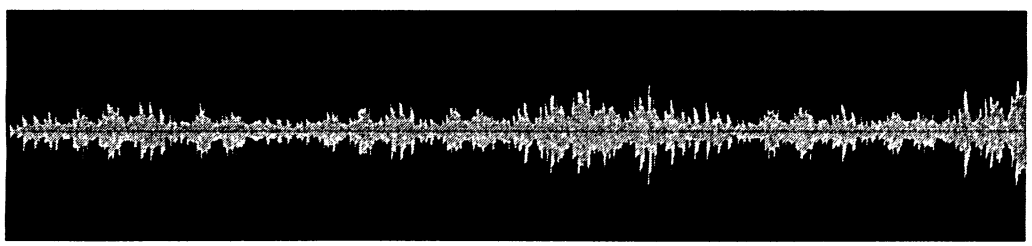


Figure 4 An example of music waveform. The first minute of Beethoven's Fur Elise.

Entropy can be defined as the amount of energy contained within a system, or as the case is with the LiveDescribe discrimination algorithm, the energy contained within an audio signal. As seen in Equation 1, an entropy algorithm measures the energy present in an audio signal. Music signals have been found to have higher entropy than signals containing human speech. I chose to use entropy because it has a high effectiveness rating, it is relatively easy to implement and the algorithm is readily available. The first step to calculating entropy is to read the raw data from the digital audio file (e.g., .wav file), and divide that data into seconds and then again into 16 ms frames (a window of data), or approximately 706 frames or samples per 1 second of audio data.

Once a window of data is obtained, the maximum and minimum value of the window is used to construct the range of the first histogram. The window is then traversed and for each occurrence of any given amplitude in the window, that corresponding amplitude in the first histogram is incremented. Once the window is traversed, we have the probabilities of each amplitude of that window in the first histogram. The second histogram size is obtained by the square root of the first histogram. Once the histogram is obtained then entropy level for the given window is obtained by summing the log of the values of the histogram. These values are stored in an array containing the entropy values of an entire second of data. The variance of the values is obtained, which represent the entropy level for that given second.

Equation 1 - Equation for calculating the histogram for the entropy algorithm (Pinquier, 2002)

$$H = \sum_{i=1}^{k} - p_i \log_2 p_i$$

with p_i = probability of event i and k = number of issues

1) A histogram is computed

Let N = number of samples contained in the considered window and S = values of amplitude taken by the signal $S = \{S_1 \ldots S_n \ldots S_N\}$

The minimal value of the histogram is defined by: $\min_h = \min(S) - \frac{\Delta}{2}$

The maximum value is defined by: $\max_h = \max(S) + \frac{\Delta}{2}$

with $\Delta = \frac{\max(S) - \min(S)}{N - 1}$

The number of stages in the histogram is defined by: $N_h = \sqrt{N}$

2) Entropy is estimated

Once the histogram is computed we have probabilities of the amplitude values. The biased estimator is computed by:

$$\hat{H}_{biased} = \frac{\sum \hat{} - h_i \log(h_i)}{N} + \log(N) + \log(\frac{\max_h - \min_h}{N_h})$$

Bias is computed by :

$$nbias = \frac{N_h - 1}{2N}$$

The estimator not biased is obtained by removing the bias:

$$\hat{H}_{unbiased} = \hat{H}_{biased} - nbias$$

This measure is used to compute the entropy modulation on one second of signal. Since we chose to calculate the entropy in 16 ms frames, we obtain

62 values of the entropy $\hat{H}$ for each second.

Another method LiveDescribe uses for discriminating speech and non-speech is zero crossing. Saunders (1996) provides one of the first examples of research using zero crossing rate (ZCR) in a speech music discrimination system. He noted that zero crossing is "one of the most indicative and robust measures to discern voiced speech" (Saunders, 1996, pg 994) and that it "provides a measure of the weighted average of the spectral energy distribution in the waveform, the spectral center of mass." (Saunders, 1996, pg 994). He describes the method for calculating zero crossing as

> "The first step is to measure the ZCR of the signal over a 2.4 second segment of the data. Next, obtain the mean ZCR and define a fixed threshold a certain number of units above and below the mean value. The last step is to count the difference between the number of points at the low end of the distribution below the lower limit and the number of samples that exceed the higher limit. If this statistic exceeds a specific threshold, the distribution outside these bounds is significantly skewed and the wave form is likely speech." (Saunders, 1996, pg 994)

The third algorithm used by LiveDescribe to discriminate speech is the root mean square. For each window described in the entropy algorithm, the amplitude value of each sample point is squared; all of the squares of the amplitude values are summed and divided by the total number of values. Once the sum for that window is obtained the square root of that sum is taken. Once RMS values for one second are obtained, the number of frames below the average amplitude value for that second is counted. This value is used to determine if that portion of audio is speech or non-speech. If the value is above the average the segment is music, if it is below it is speech.

Pinquier showed that when an entropy algorithm (similar to the one used in LiveDescribe) was combined with other discrimination algorithms, such as 4hz modulation and segment analysis, a speech / music discrimination system with an accuracy rate better than 90% could be realized. RMS and zero crossing algorithms were chosen instead of 4hz and segment analysis due to availability of detailed algorithm

definitions and a desire to explore a multitude of solutions. 4hz modulation and segment analysis may be implemented into future versions of LiveDescribe. The effectiveness of the entropy/zero crossing/RMS combination has not been tested rigorously. Future research will look at the effectiveness of various combinations of speech discrimination algorithms with numerous genres and types of programming.

Determining whether the algorithm combination produced an accurate valuation of the audio signal was not within the scope of this thesis, however, I wanted to ensure that a majority of the possible description spaces were detected. The conventional approach in the description process is for a human to analyze the audio track and determine the spaces and timing for possible description. A human rater thus manually defined the description spaces in *The Daily Show* episode used in this thesis, which were then compared with the ones that were automatically detected by LiveDescribe. Using a one second tolerance window, the human rater and automatic discrimination system were found to be in agreement 89% (1129 agreements, 131 disagreements) of the time. Although the LiveDescribe discrimination algorithm has not yet been tested on a wide range of content, this early positive result using this specific content is promising.

3.4 Conceptual Design of LiveDescribe Source Code

LiveDescribe was designed and coded using an engine-based conceptual framework in order to keep similarly themed functions together in relatively isolated modules. Although I attempted to adhere to proper object-orientation design concepts some concessions were made in the interest of practicality and time limitations. Similar functions of LiveDescribe are grouped into conceptual engine classes from which these

functions can be called. LiveDescribe currently uses five conceptual engines including capture, discrimination, editing, video and wave file. These engines handle video and audio capturing, speech / non-speech discrimination, audio editing, video playback and wave file display respectively. In addition, functionality that does not fit into these conceptual engines are placed into either helper classes or graphics classes. The graphic classes are responsible for generating and displaying the description track, the entropy graph and individual graphical waveform representation of descriptions respectively, while helper classes contain any remaining miscellaneous functionality.

3.5 Video Description Creation Process/Workflow Overview

As seen in Figure 5, LiveDescribe attempts to mediate an efficient work flow for creating and publishing video description by separating the tedious physical requirement of determining when and where to describe within a specific piece of content from the more difficult cognitive task of deciding what visual information to describe and how to describe it. The original audio and video data is inputted into the automated speech/music discrimination algorithms, which then presents the video describer with enhanced audio and video data containing information about description spaces and their position and duration. The human describer can then concentrate on the decisions necessary to produce description without the tedious task of finding and measuring points in the original audio where description can be inserted. The option to modify the spaces and time is available to users as is the option to use non-detected spaces. However, making these modifications are exceptions rather than common tasks. The output of this system is

the combination of the original audio/video tracks with the describer created description track, appropriately timed and adjusted for volume levels

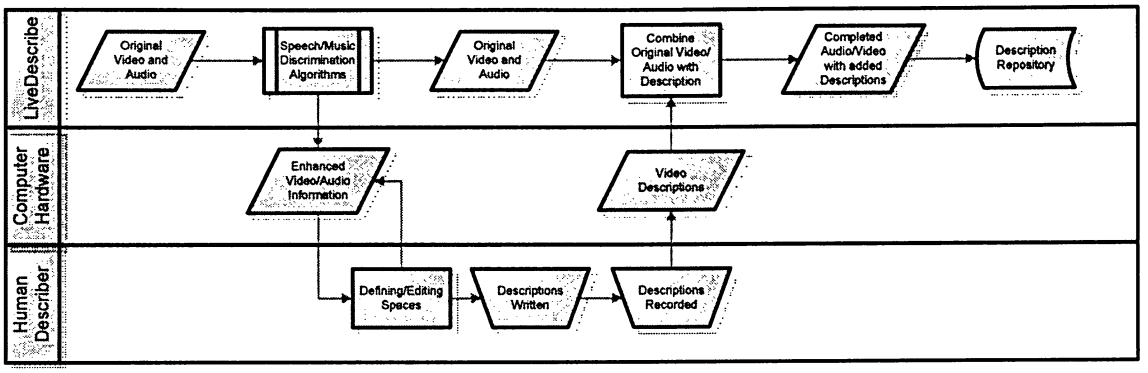


Figure 5 - A process flow diagram illustrating the work flow for creating video description using LiveDescribe

In comparison, Figure 6 illustrates the process flow for creating description manually without the assistance of computer hardware or the LiveDescribe software. As seen in the figure, the describer is required to carry out all of the tasks including manually locating description opportunities (identifying spaces and their duration), writing the descriptions in a separate system, recording the description using yet another system and then adjusting the recorded description to fit within the desired spaces. This approach to description means the describer is spending not only more time in general on the overall task of creating description but also more time on laborious subtasks such as finding spaces and adjusting description and likely less time on high level cognitive tasks such as writing and recording description.

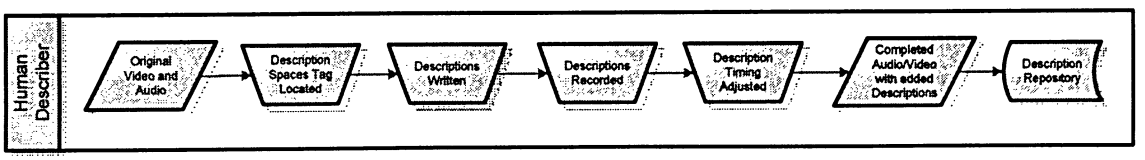


Figure 6 – The work flow process diagram for manual description creation

4. User evaluation

4.1 Research Questions

The primary objective of this research is to determine the feasibility of amateur describers creating usable video description using LiveDescribe. Traditional post production description is currently created by "professional" describers who receive training, and usually possess appropriate skill sets, such as a good speaking voice, an extensive vocabulary or performance arts training, required to produce high-quality description. Often these descriptions are created in a dedicated professional studio environment using professional quality audio recording and editing equipment that is expensive to purchase or rent, and maintain. One potential reason why more description is not produced by non-professionals is the lack of availability and cost of production equipment and recording studios. A second, perhaps more important, reason is that there is no simple method for creating descriptions that would allow non-professionals to record, edit, mix and ultimately deploy video description for television and film.

A second main objective of this research is to begin to understand the characteristics and attributes of a successful and unsuccessful non-professional describer. Knowing what makes a good or poor describer will help in the future creation of tutorials, guidelines or other documents designed to train new and existing amateur describers as well as the future design and creation of video description software or hardware tools.

Three specific research questions that follow from these objectives and are addressed in this thesis are:

1. Are amateur describers able to create video description with acceptable quality levels?

2. Is the process of creating description, mediated through LiveDescribe, viable and practical?
3. Is LiveDescribe a viable tool for assisting in the creation of video description by amateur describers?

4.2 Method

Two phases of exploratory studies were designed to begin to answer these research questions. During the first phase, description by non-professional describers was created in order to be reviewed during phase II. In phase II, the descriptions generated in phase I were reviewed using four main factors including overall quality, vocabulary and audio quality. Phase I studies were conducted in the usability lab at the Center for Learning Technology at the Ted Rogers School of Management while phase II was conducted online. The research was approved by the Ryerson Ethics Board. The letter of approval can be seen in appendix G.

An Internet-based research method was chosen because the access to worldwide participant base increasing the number of possible participants substantively over the number that could be recruited locally, it is relatively simple to implement and there is evidence that it is an effective and efficient methodology. Hammontree (1994) suggested that with advances in group collaboration technologies that remote usability methods should be considered. At the time he predicted that because of increases in network bandwidth, the majority of usability testing could be performed remotely. Dahan (1998) showed that an online 3D virtual product demonstration captured similar user reactions as a face-to-face demonstration using a physical prototype. This suggests that using online

customer or audience satisfaction testing shows promise and that applying such a method to satisfaction testing for video description is possible.

One of the important variables to be measured in my studies is that of "acceptable quality." Jumisko-Pyykko (2006) suggests that audiovisual quality is a combination of perceived and produced or technical quality. In a study of mobile phone video quality, using an 11-point unlabeled Likert scale, she found that the most noticeable differences of perceived quality appears at the threshold between acceptable and unacceptable levels. This means that after a certain level of quality, increases did not result in a correlated increase in quality perception. She also found that users perceived the quality level as low if the conveyed information was irrelevant or of low quality even when the production quality was high.

International Telecommunication Union (ITU) also provides extensive guidelines and recommendations for the subjective quality assessment of audiovisual content. They suggest that measuring perceived quality requires the use of a subjective scaling method, specifically a 5-point Likert scale with the categories bad, poor, fair, good and excellent. The ITU guidelines, the results of the Jumisko-Pyykko study and recommendations from other researchers such as Preece (1994) indicate that it is important to measure perceived quality using a subjective scale when evaluating viewer reactions rather than assessing the technical quality of audiovisual content including video description. In the second phase of my research, perceived quality was thus evaluated using 5-point Likert scales.

Preece (1994) also suggests that video logging provides a preferred alternative to direct observation because it can reduce the Hawthorn effect. For this reason I chose to capture video data for later analysis instead of observing directly. Collecting video data

also meant that I could review recorded events many times over and perform detailed analyses.

Phase 1

During the first phase, 12 participants, 5 female and 7 male, created video description for an entire 20 minute episode of *The Daily Show* with Jon Stewart. This episode was downloaded from the internet in DivX AVI format. Eight of participants were in the 18-29 range, two from each of the 30-39 and 40-49 categories and one from 50-59 age range. The education background of the participant group was almost equally distributed with similar numbers in all education categories from high school to graduate studies with three having a high school diploma, two with a college diploma, five with a university degree, and two with a graduate degree. Only one of the participants had ever created description before, but not regularly. Four participants had heard of description prior to this experiment, but had never created it before, while the rest had never heard of and consequently had never created description.

Upon arrival to the test facility, participants were given a short 17 question pre-study questionnaire (see Appendix A), which was designed to capture demographic information as well as any previous experience with video description. They were also asked for their familiarity and level of appreciation for *The Daily Show*, and their level of computer proficiency. Once this pre-study questionnaire (fifteen forced-choice Likert scale questions and two open ended questions) was completed, participants were then informed of the purpose of this research and given a brief overview of the tasks in which they would be involved over the course of the study.

Next, training was provided for video description and how to use LiveDescribe to create and record descriptions. Video description training was very brief, lasting about 15 minutes, and consisted of informing the participants of some of the conventions of description, such as attempting to keep descriptions within pauses of dialog, and avoiding describing visual elements that also have an audio cue (telephone ringing) (The Independent Television Commission, 2000). However, stylistic conventions were not included as participants were specifically told to adopt their own personal style.

After this brief discussion and after answering any initial questions, two examples of described versions of clips of the cartoon Odd Job Jack of were shown to the participants. The first clip contained conventional descriptions using the third person, omniscient, mono-tone based narrative form, while the second clip contained description using first-person narration, provided by the main character Jack. These two contrasting examples served to illustrate to the participant that although there are general conventions for creating quality video description, it can also be a creative or artistic process. Once these clips were reviewed and remaining questions about description were answered, the next training session involving an orientation to LiveDescribe began.

Training in LiveDescribe was also brief, and consisted mostly of an introduction to the main functionality of LiveDescribe. Each of the elements/functions of LiveDescribe was introduced and then the steps for performing the tasks associated with that function were demonstrated. After this brief overview, participants were asked to complete three simple description tasks to assure they had learned the basic functionality of LiveDescribe. First they were asked to record a description over any part of the video, second, they recorded a pre-typed test description and finally they wrote and recorded

their own description. If they were able to complete these tasks without asking for assistance they passed the test. If they had additional questions or were unable to complete the tasks additional instruction was given until they were able to. All participants were able to pass this test within a few minutes.

Once these training tasks were completed satisfactorily participants were asked to complete the task of creating description of *The Daily Show*. They were instructed that I could provide assistance only in the event of an unexpected technical problem, such as a software crash, or audio recording problem, but that any issues concerning description would have to be decided on their own.

After the participant had completed the task of creating description, a post study questionnaire (ten forced-choice Likert scale questions and three open ended questions) designed to capture the participant's experience of describing was administered. This 13 question survey asked about usability, ease of use and usefulness of both LiveDescribe and more generally the task of video description (see Appendix C). In addition the entire training and study session was video recorded using a Sony DC-SR100 30GB hard drive camera for analysis.

Phase II

In phase II, 76 blind, low vision and sighted participants reviewed the descriptions created in phase I and commented on various aspects of the description such as overall quality, vocabulary level and style of delivery. Six participants were sighted, 25 were low vision and 45 were totally blind. For each describer from phase 1, an audio only version of the Daily Show including the created descriptions was generated and divided into 5

separate segments in order to ensure the maximum number of describers could be reviewed while still maintaining the integrity of the episode.

Each segment break was made at an appropriate point, such as during a commercial break, or at the beginning of a new sketch or show segment. Each segment was 3-5 minutes in length. The audio-only segments were then presented to each participant in chronological order, but with a randomly selected describer. Each clip was compressed to MP3 format and stored on a web server for public access.

Participants were asked to complete a total of six surveys; the first consisted of eleven force-choice Likert scale questions and two open ended questions that collected demographic information such as vision status, age, employment status, and familiarity and level of appreciation of *The Daily Show* (see Appendix D). The second set of surveys consisted of the same survey administered five times; one for each of the five clips. This survey consisted of eleven forced-choice Likert scale questions and two open ended questions that asked about the overall quality, audio quality and style the description clip they had just reviewed (see Appendix F).

4.3 Experimental Setup and Data Collection

Pre and post questionnaires for phase I were administered in person, but performed online through a website using PHP scripting and MYSQL database technology. Questionnaire data was entered directly into the computer and stored in a relational database.

Participants were asked to perform their describing tasks in a usability laboratory that consisted of an office chair, a desk with a Dell Precision Desktop PC, a 22 inch wide

screen LCD monitor and a simple low cost headset and microphone combination. A video camera was setup behind the participants and captured the screen.

The pre-study and post-viewing questionnaires of phase II were administered online. This allowed for a large number of participants from anywhere in the world to participate using their own computers and Internet connection.

4.4 Data Analysis

Phase I

Questionnaire data was collected from a total of 12 phase I participants. Likert scale responses from the questionnaire were coded using 5 as the highest positive value (very easy or very useful) and using 1 as the lowest positive value (very difficult or very useless). A chi-square analysis was then applied to the responses in the questionnaire to examine the comparison with an expected distribution of equal numbers in each answer category.

Thematic video analysis was conducted on the video data recorded during phase I. As seen in Table 1 data was collected for two categories: description creation and errors. There were three subcategories of the description creation category: text with space, no text with space, no text and no space and three subcategories for errors: navigation error, recording error and text saving error. A reliability analysis was conducted using two independent raters on a fifteen minute segment of video. Using the results from both independent raters Cohen's kappa coefficient was 0.81 and Pearson's Rho was 0.97

which indicates that the inter-rater reliability was excellent. All subsequent analyses were carried out by a single evaluator.

Table 1 - Thematic analysis catgory and modifier definitions

Category	Modifier	Definition
Descriptions	Text With Space	A description was created by first typing text into the text box, clicking the timed record button, and recording the description using the text highlighting tool.
	Space No Text	A description was created using a manually or automatically defined space as a guide, but no text was used.
	No Text No Space	A description was created without the aid of a defined space or text.
Errors	Navigation	The user attempt to navigate to another position in the video but instead activated another feature, or navigated to an undesired position.
	Recording	Users would attempt to record in a spot where description already existed, or tried to record but instead activated another function like play or fast forward.
	Text Saving	Users entered text and, without saving it, attempted to record the description. This results in the entered text being lost.

Phase II

Questionnaire data was collected from 76 participants who gave 287 clip reviews in total. Although some participants did not finish the entire survey, the clip reviews they did perform were used in final analysis. Thirty-six participants completed all five clip reviews surveys, 12 participants completed four reviews, 10 participants completed three reviews, 7 participants completed two reviews and 10 participants completed only one review. One way ANOVA's were used to examine the relationships between reviewer perception of quality (Overall Quality, More Information, Listen to More etc) and various characteristics of the reviewer or the description itself determined through the video analysis and questionnaire data. A Bonferonni adjusted significance level of 0.01 was obtained by dividing 0.05 by 5, the number of ANOVA tests conducted on a single set of

data. This was done to safeguard against the increased probability of Type I errors resulting from multiple tests of statistical significance on the same data set. Although variables are reported with a $p < 0.05$ significance level to illustrate trends which may warrant further study, variables which passed the Bonferonni adjusted significance level of $p < 0.01$ have been highlighted in the text in a bold font.

5. Results

5.1 Phase I

A chi-square analysis revealed significance for ten of the phase 1 post-study questions which are listed below in Table 2 and Table 3. All values are reported to $p<0.05$ level. A total of 12 participants responded to each question. This analysis was performed in order to compare the study responses to an expected equal distribution of responses in each category.

Table 2 - Significant Ease of Use Variables

Variable	χ-Square	Mean	Standard Deviation
Leaning to use LiveDescribe	15.85	4.00	1.28
Understanding the video	16.62	3.92	0.64
Using the computer	21.23	4.62	0.51
Editing Description Boundaries	14.01	3.53	1.05
Learning to use Software	13.00	4.00	0.85

Ten of the 12 participants reported Learning to Use LiveDescribe was "easy" or "very easy", while the remaining two reported "not easy nor difficult" and "difficult". All participants reported that Using the Computer was "easy" or "very easy". Ten of the participants reported that Understanding the Video was "easy" or "very easy", two said it was "neither easy nor difficult". One participant reported Editing Description Boundaries as "very easy", while seven report it as "easy". Two reported Editing Description Boundaries was "neither easy nor difficult" task and the final two reported it as hard and very hard. Nine participants reported Learning to Use Software as "easy" or "very easy"

while the final two reported "neither easy nor difficult" and "difficult". Table 2 shows all the significant results of the ease of use questions.

Table 3 shows the results a chi-square analysis on the usefulness questions. All values are reported to $p<0.05$ level. Ten of 12 participants said the Graphical Timeline and Automatic Space Detection was "useful" or "very useful", two said the Graphical Timeline was "neither useful nor useless", and two said the Automatic Space Detection was "useless". Ten participants said the Timer for Recording was "very useful" or "useful", while two reported it was "neither useful nor useless". Editing Description functions was reported as being "very useful" or "useful" by eight participants, "neither useful nor useless" by 3 participants and "useless" by one participant. Five participants said the Running List of Description was "useful" while the remaining reported it as "neither useful nor useless".

Table 3 - Significant Usefulness Variables

Variable	Chi-Square	Mean	Standard Deviation
Graphical Timeline	12.00	4.31	0.75
Automatic Space Detection	16.62	3.92	0.95
Timer for Recording	16.12	4.08	0.64
Editing Description Functions	11.23	3.77	0.83
Running List of Spaces	15.85	3.31	0.63

Thematic analysis was conducted on the video data recorded during Phase I. Data concerning the number of total descriptions made and whether or not description spaces or text tools were used to assist in the creation of those description, as well as errors made by the describers were collected. Errors included text saving errors, where a describer would enter text, but fail to save it before attempting to record their description; navigation errors, where describers attempted to navigate to a part of the video but

instead activated another function, or navigated to the incorrect place. In addition, variables such as total descriptions created, total length of descriptions, and average length per description were counted.

Table 4 shows the results of the video analysis conducted on the video data collected from Phase I participants. The first column contains the describer identification number. The second column *total* representing the total number of final descriptions the describer ended with, while the third represents the time in seconds of all of the final descriptions. The third column *mean*, is the average length of the final descriptions made. *Total Recorded* represents the total number of descriptions the participant made during the experiment, including those that were made, but then deleted and not included in the final version. The *count/made* ratio is the total number of final descriptions divided by the total made overall. *Wrong* is the total number of conceptual errors the participant made during the experiment. These include such actions as failing to save description text before recording or navigation errors. The last column represents the total time each participant took to complete their description task.

Table 4 - Video analysis data sorted by describer identification number.

Participant ID	Total	Total Length (sec)	Mean (sec)	Total Recorded	Count / Made ratio	Errors	Duration(Sec)
1	26	67.91	2.61	31	0.84	0	2417.89
2	14	75.71	5.41	30	0.47	4	2780.36
3	28	246.37	8.80	104	0.27	25	4647.28
4	18	53.04	2.95	41	0.44	2	3920.22
5	23	118.19	5.14	34	0.68	6	5546.91
6	57	189.05	3.32	N/A	N/A	N/A	N/A
7	13	40.87	3.14	28	0.46	16	4042.33
8	12	34.72	2.89	15	0.80	15	3185.54
9	35	122.79	3.51	83	0.42	9	8845.66
10	25	113.96	4.56	58	0.43	16	5241.33
11	58	468.80	7.81	60	0.97	5	3142.87
12	36	165.74	4.60	N/A	N/A	N/A	N/A

Table 5 shows a more detailed breakdown of how each describer used LiveDescribe to complete their description task. When describers used a space offered by LiveDescribe and the text highlighting tool their description was counted as a "Text/Space" type of description. Similarly if describers used a space for guidance but did not enter text, it was marked as a "No Text/Space" and finally if a describer created a description without the use of a space as a guide or entering text it was entered as "No Space/No Text". The situation "No Space / Text" could not exist since a space is required to use text. Unfortunately due to data loss, data from describers D6 and D12 were not available.

Table 5 - Number of descriptions made be each describer in the description categories.

Participant ID	Text / Space	No Text / Space	No Text / No Space
1	2	29	0
2	1	29	20
3	9	74	0
4	0	40	1
5	2	10	22
6	N/A	N/A	N/A
7	28	0	0
8	14	1	0
9	78	5	0
10	34	30	10
11	0	15	43
12	N/A	N/A	N/A

5.2 Phase II

This phase of the research was focused on assessing the viewer perspective on the results of amateur descriptions. Assessing viewer reactions and opinions of the descriptions as created by various describers and what characteristics of the describer or description resulted in those opinions were the main objectives of this phase. The focus of data analysis was to find relationships/differences between describers or description characteristics and the perception of quality by reviewers.

A one way ANOVA was conducted on the phase 2 survey data to assess differences of reviewer opinion between describers from Phase I. The ANOVA analysis requires that a random sample be used and that there is homogeneity of variance in the groups. Although every effort was made to assure all samples where as random as possible, overall quality, more info and more entertainment did not pass a test for homogeneity of variance when all participants were included. This assumption can be relaxed however if the sample size between groups are similar, which is the case with this data. Describe D12 received the lowest amount of reviews (19) with Describer D8 receiving the most reviews (34) while most describers received between 23 and 27 reviews. When the 84 reviews completed by participants who had no experience with description were excluded from the sample, all variables passed the homogeneous of variance assumption. Nine of this excluded group of reviews was completed by blind participants, 55 were completed by participants with low vision and 20 were completed by sighted participants. This suggests that most of the blind participants had a least some experience with description.

When including all participant reviews in the data set, a significant difference between describers for all variables except Audio Quality was found. Table 6 shows the ANOVA results for all significant variables to a level of $p<0.05$.

Table 6 Summation of results for all significant variables to a level of $p<0.05$ as grouped by describer

		df	F	Sig.
overallquality	Between	11	4.377	**.000**
	Within	274		
	Total	285		
morehumour	Between	11	3.755	**.000**
	Within	273		
	Total	284		
moreinfo	Between	11	1.889	.041
	Within	271		
	Total	282		
moreentertainment	Between	11	2.111	.020
	Within	257		
	Total	268		
vocablevel	Between	11	4.009	**.000**
	Within	270		
	Total	281		
styleofdelivery	Between	11	2.934	**.001**
	Within	270		
	Total	281		
listentomore	Between	11	3.383	**.000**
	Within	273		
	Total	284		
comparedtopro	Between	11	3.001	**.001**
	Within	262		
	Total	273		

Three describers were rated as most preferred, 3 describers were rated as least preferred and 6 describers were grouped in the centre of the preference scale. Describer D12, D10, and D9, appear to be the preferred describers for Overall Quality ($M = 3.5$, $SD = 1.34$; $M = 3.48$, $SD = 1.38$; $M = 3.37$, $SD = 3.37$ respectively) where Overall Quality was rated from 1(very poor) to 5 (very good) on a Likert scale. Fifteen of 19 reviewers of

Describer D12, 20 of the 25 reviewers of Describer D10 and 20 of the 25 reviewers of Describer D9 said that the Overall Quality was okay, good or very good.

Describer D11 ($M = 2.26$, $SD = 1.60$), 8 ($M = 2.20$, $SD = .977$) and D4 ($M = 2.1667$, $SD = 1.0072$) all achieved significantly lower mean values for Overall Quality and appear to be the least preferred describers. Fifteen of the 23 reviewers of Describer D1, 21 of the 34 reviewers of Describer D8 and 16 of the 24 reviewers of Describer D4 all reported the Overall Quality to be poor or very poor. I used overall quality as the primary measure of description quality throughout data analysis because it provided a summary judgment. Other questions such as style and language used and information conveying ability, which all contribute to overall quality, are related to specific elements of quality and showed similar groupings as Overall Quality in means and standard deviations. Figure 7 shows the mean overall quality rating and standard deviation for each describer.

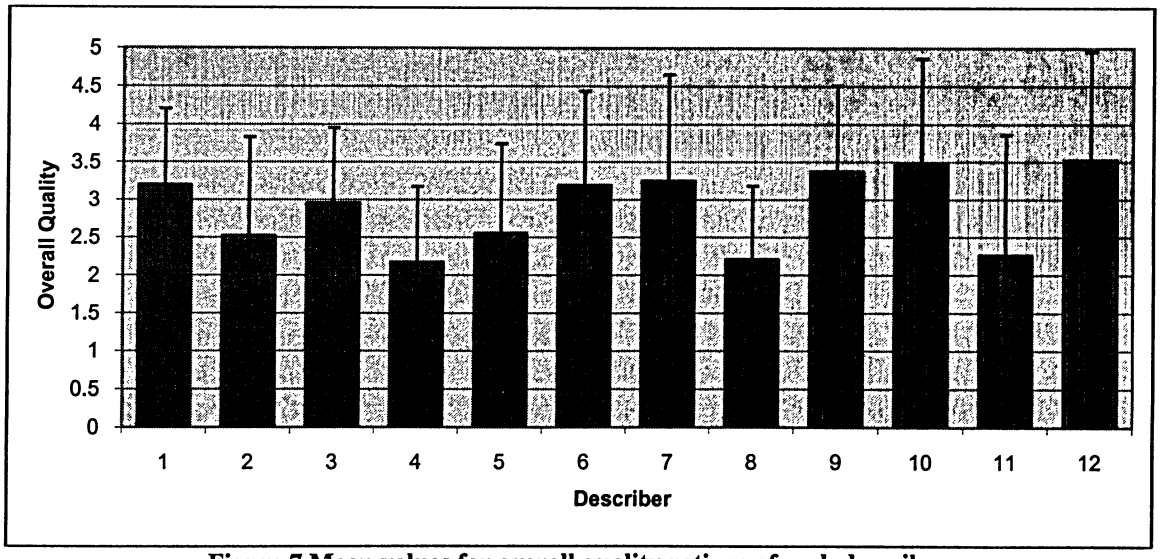


Figure 7 Mean values for overall quality ratings of each describer

As seen in figure 8 and further illustrating the effect of describer on quality rating are the results seen from the compared to professional question. Similarly to the overall

quality results describers D12 ($M = 2.63$, $SD = 1.30$), D10 ($M = 2.25$, $SD = 1.07$) and D9 ($M = 2.43$, $SD = 1.16$) all achieved the highest rating for compared to professional, 5 being "a lot better" than a professional,1 being "a lot worse" and 3 being "as good". This result suggests that the quality level of these top describers is very close to that of professionally produced description.

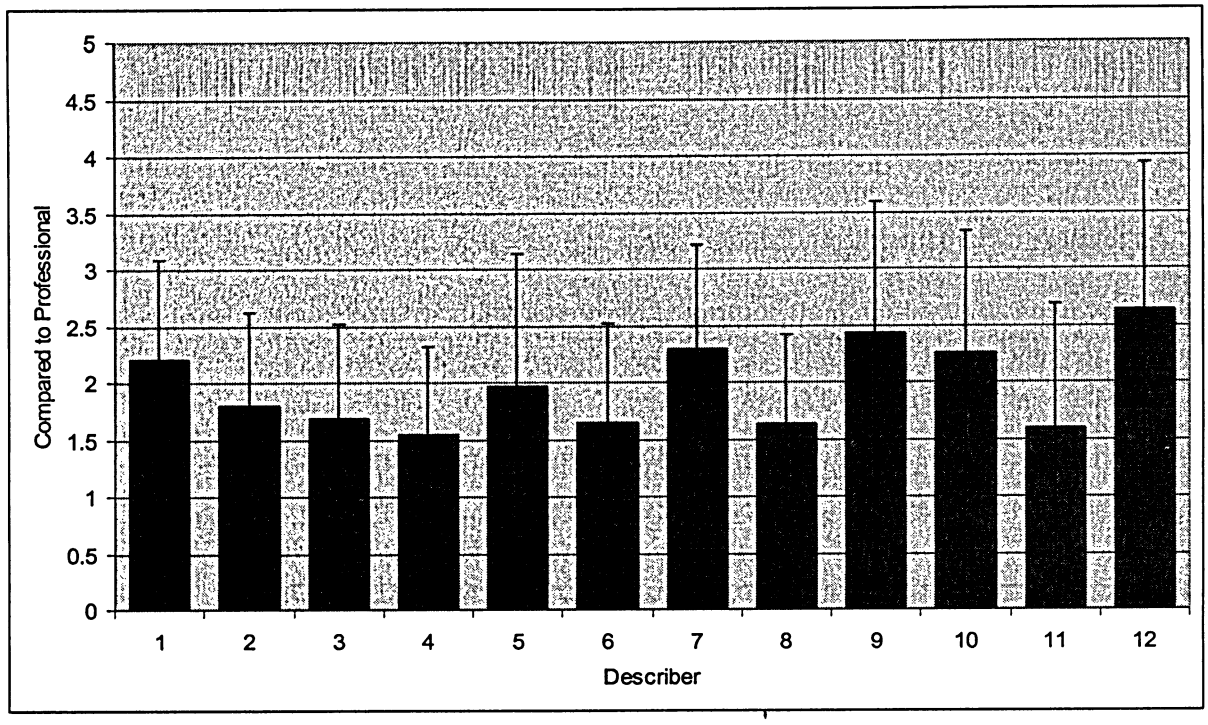


Figure 8 - Mean values for compared to professional ratings of each describer

When participants without experience with description were removed from the dataset, the remaining participants have professional description as a point of reference for comparison of quality with the amateur description in this study. Figure 9 shows the mean values of overall quality for each describer excluding those with no description experience. Describer D12's overall quality rating increases even further ($M = 3.88$, $SD = 1.13$) while describe D11 drops lower ($M = 1.77$, $SD = 1.30$) and describer D6 and D7 move from the medium rated category in the good category. This suggests that

experience with description may play a role in quality perception but further investigation of the extent of the impact of experience remains.

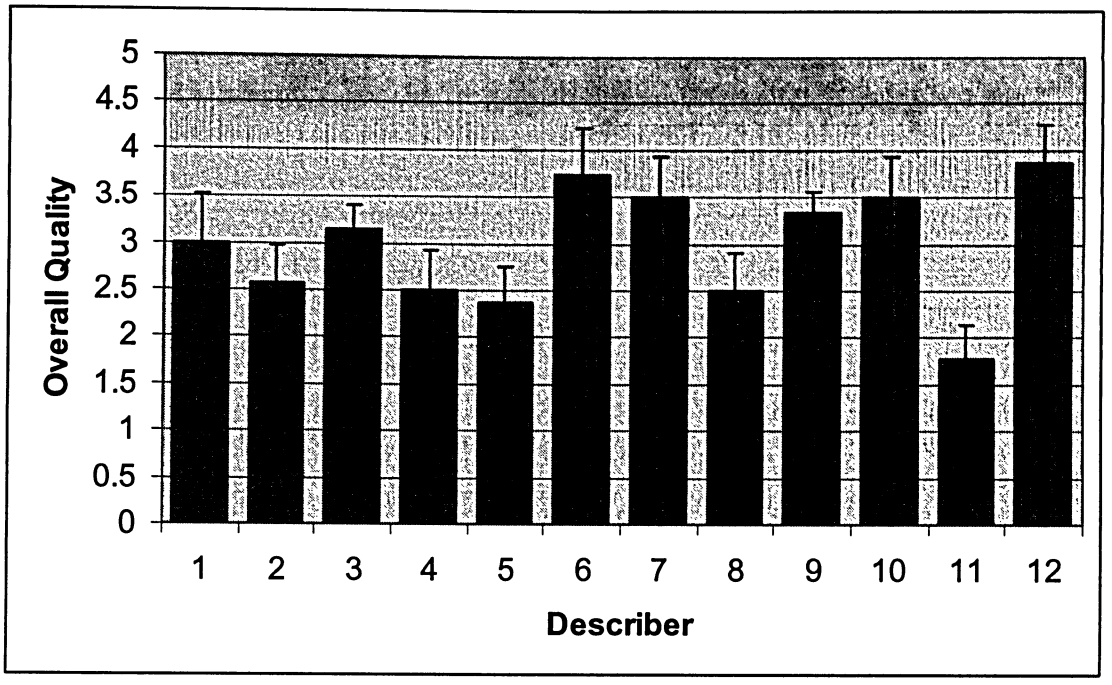


Figure 9 - Mean values for overall quality of each describer, not including reviewers who lack experience with professional description

Tukey HSD post-hoc analysis showed that the differences are primarily between the worst and best describer. As seen in table 7 there is significance between mean values for overall quality between describer D12 (the highest rated describer) and D11, D8 and D4 (the lowest rated describers).

Table 7 - A Tukey HSD Post-Hoc analysis showing the significant differences for overall quality between describer 12 and other describers

Dependent Variable	Describer I	Describer J	Mean Difference (I-J)	Sig.
Overall quality	12	4	1.35965(*)	.018
		8	1.32043(*)	.011
		11	1.26545(*)	.045

Table 8 shows significance difference for overall quality between describer D11, the poorest rated describer, and describer D12 and describer D10, two of the three top rated describers.

Table 8 A Tukey HSD Post-Hoc analysis showing the significant differences for overall quality between describer 11 and other describers

Dependent Variable	Describer I	Describer J	Mean Difference (I-J)	Sig.
Overall quality	11	10	-1.21913(*)	.031
		12	-1.26545(*)	.045

A one way ANOVA was conducted to determine the effect of impairment groups (blind, low vision and sighted) on describer reviews. There was a significant difference between blind, low vision and sighted viewers for all review variables except for More Information, Vocabulary Level and Style of Delivery.

Table 9 Summation of results for all significant variables to a level of p<0.05

		df	F	Sig.
overallquality	Between	2	3.081	.047
	Within	283		
	Total	285		
morehumour	Between	2	5.660	**.004**
	Within	282		
	Total	284		
moreentertainment	Between	2	4.820	**.009**
	Within	266		
	Total	268		
audioquality	Between	2	6.377	**.002**
	Within	281		
	Total	283		
listentomore	Between	2	4.033	.019
	Within	282		
	Total	284		
comparedtopro	Between	2	6.928	**.001**
	Within	271		
	Total	273		

Description appears to be more harshly judged as the vision status of the reviewer increases. For example for overall quality, sighted users reported the lowest mean (M = 2.37, SD = 1.09), while low vision reviewers reported a slightly higher mean (M = 2.70, SD = 1.35) and finally blind users reported the highest values for overall quality (M = 2.98, SD = 1.28) (see Figure 9). A Tukey-HSD post-hoc analysis however revealed no significance between groups for overall quality and this result could be due to chance. Further study is needed to elaborate on this finding.

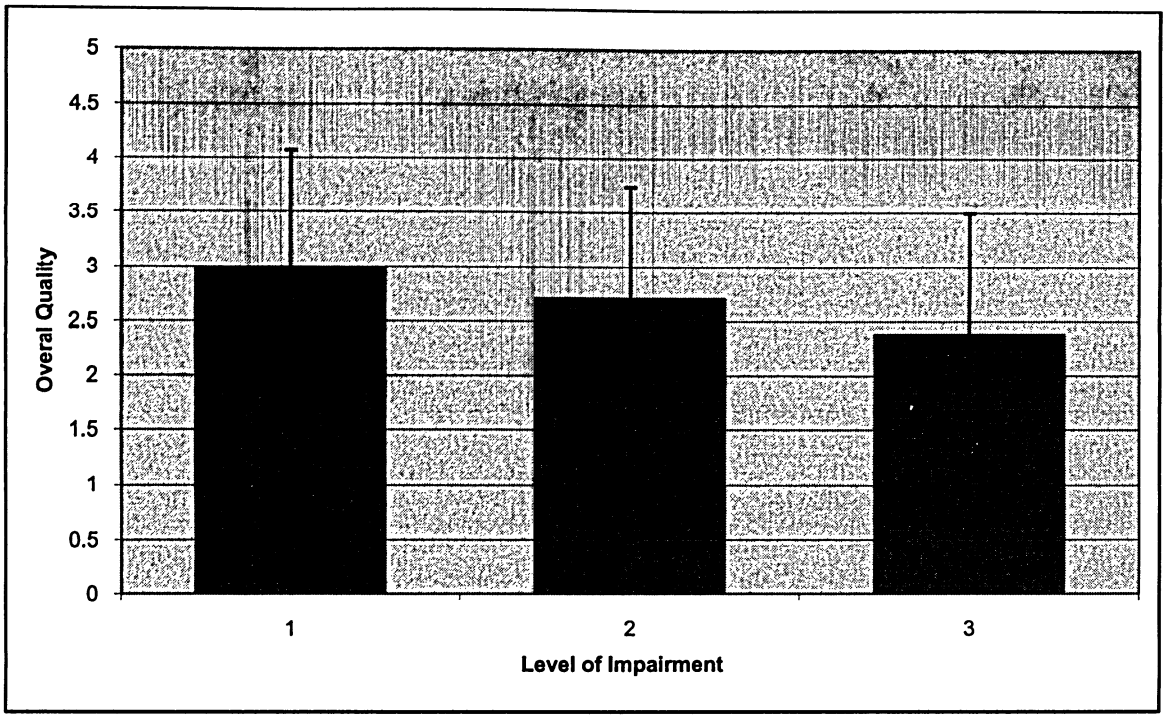


Figure 9 - Mean overall quality grouped by level of vision impairment

A one way ANOVA was conducted to determine significant differences of description opinion between each of the five description clips. There were significant differences reported between clips for all reviewer variables except Compared to Pro (see Table 10 for ANOVA results).

Table 10 - Significant differences reported between clips for all reviewer variables except Compare to Pro

		df	F	Sig.
overallquality	Between	4	5.292	**.000**
	Within	281		
	Total	285		
morehumour	Between	4	2.809	.026
	Within	280		
	Total	284		
moreinfo	Between	4	7.440	**.000**
	Within	278		
	Total	282		
moreentertainment	Between	4	3.370	**.010**
	Within	264		
	Total	268		
vocablevel	Between	4	2.614	.036
	Within	277		
	Total	281		
styleofdelivery	Between	4	3.003	.019
	Within	277		
	Total	281		
audioquality	Between	4	4.090	**.003**
	Within	279		
	Total	283		
listentomore	Between	4	3.596	**.007**
	Within	280		
	Total	284		

Post hoc tests (see Table 11) reveal a significant difference between clip 2 and clips 1, 4 and 5. For example, Clip 2 had a significantly lower review for Overall Quality ($M = 2.31$, $SD = 1.11$) compared to clip 1 ($M = 3.06$, $SD = 1.14$), 3 ($M = 2.62$, $SD = 1.48$), 4 ($M = 3.02$, $SD = 1.45$) and 5 ($M = 3.26$, $SD = 1.17$).

Table 11 - A Tukey HSD post-hoc analysis showing the significant differences between description clip 2 and clip 1, 4 and 5

Dependent Variable	(I) clipnumber	(J) clipnumber	Mean Difference (I-J)	Sig.
overallquality	1.00	2.00	.75244(*)	.005
	2.00	1.00	-.75244(*)	.005
		4.00	-.70924(*)	.032
		5.00	-.95417(*)	.001
		2.00	.70924(*)	.032
		2.00	.95417(*)	.001

Figure 10 shows the mean values for Overall Quality as grouped by clip number. As can be seen, clip 2 has the lowest mean value while clip 5 has the highest. This trend could suggest a chronological effect on the perception of quality. In addition, a characteristic of clip 2, such as not actually requiring any description, could have an effect on the perception of quality as well. It seems from the data that describers did not produce as much description for clip 2 as the other clips, however without further study regarding description qualities and quantities, it is difficult to ascertain whether this was the factor that contributed most to the result. Other factors such as appreciation for clip content, and ordering or a learning effect could also have affected this result and further study is required to determine which specific factors contribute most to people's perception of quality.

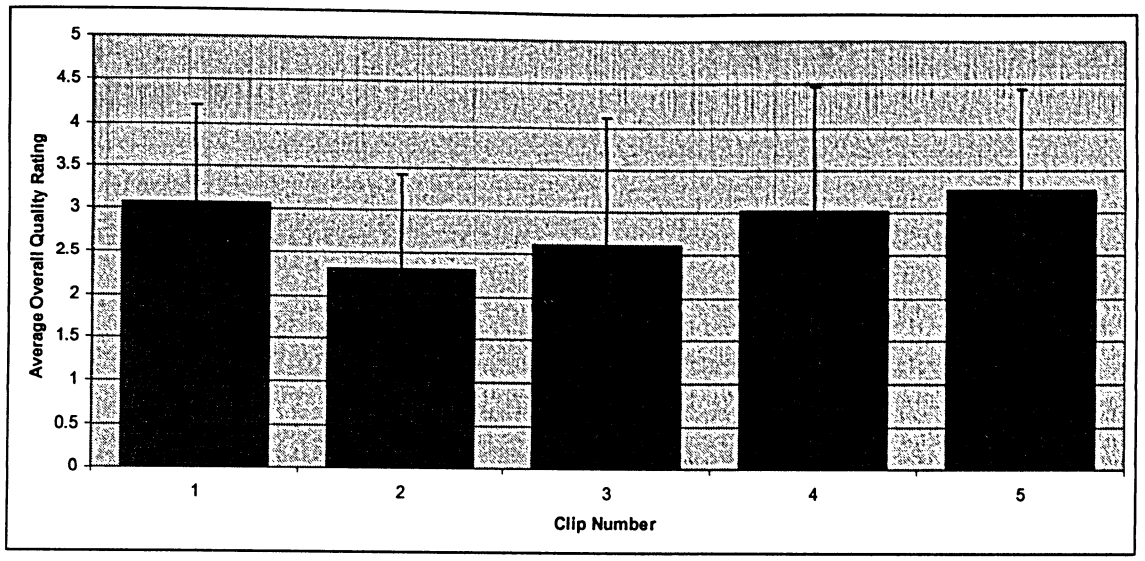


Figure 10 Mean and standard deviation results for Overall Quality Rating grouped by clip number

A relationship between the demographic data of the respondents and their opinions of the various description variables was discovered. A one way ANOVA was conducted to determine significant differences between reported cinema attendance frequency and reviewer opinion. There were significant differences reported between groups of cinema frequency (never, once a year, once a month, once a week, more than once a week) for all reviewer variables except Vocabulary Level, Style of Delivery and Compare to Pro.

Table 12 - Significant differences reported between cinema attendance frequency for all reviewer variables except Compare to Pro

		df	F	Sig.
overallquality	Between Groups	4	3.569	**.007**
	Within Groups	281		
	Total	285		
morehumour	Between Groups	4	3.354	.011
	Within Groups	280		
	Total	284		
moreinfo	Between Groups	4	5.409	**.000**
	Within Groups	278		
	Total	282		
moreentertainment	Between Groups	4	3.726	**.006**
	Within Groups	264		
	Total	268		
audioquality	Between Groups	4	5.573	**.000**
	Within Groups	279		
	Total	283		

Post hoc tests reveal a significant difference between reviewers who attend the cinema more than once per week and reviewers who attend much less or never. For example, Overall Quality for reviewers who attend the cinema more than once per week report the lowest result ($M = 1.20$, $SD = .44721$), while those that attend the cinema once per month ($M = 2.80$, $SD = 1.29$), or never ($M = 2.93$, $SD = 1.28$), rated Over Quality as higher. This trend continues across all of the reviewer variables. As discovered in a post-hoc Tukey HSD revealed most of the difference is from group 5, those that attend the cinema more than once a week. A limitation of this finding is that only 5 participants reported going to the cinema more than once a week. Further study will be needed to confirm this finding.

Table 13 A Tukey HSD post-hoc analysis showing the significant differences between reviewers cinema attendance level with overall quality as the dependant variable

(I) movies	(J) movies	Mean Difference (I-J)	Sig.
1.00	5.00	1.73478(*)	.034
2.00	5.00	1.60702(*)	.049
3.00	5.00	1.79083(*)	.020
5.00	1.00	-1.73478(*)	.034
	2.00	-1.60702(*)	.049
	3.00	-1.79083(*)	.020

Figure 11 shows the mean values of overall quality as grouped by review cinema attendance level. Five means the reviewer attends the cinema once a week or more, while one means the reviewer never attends.

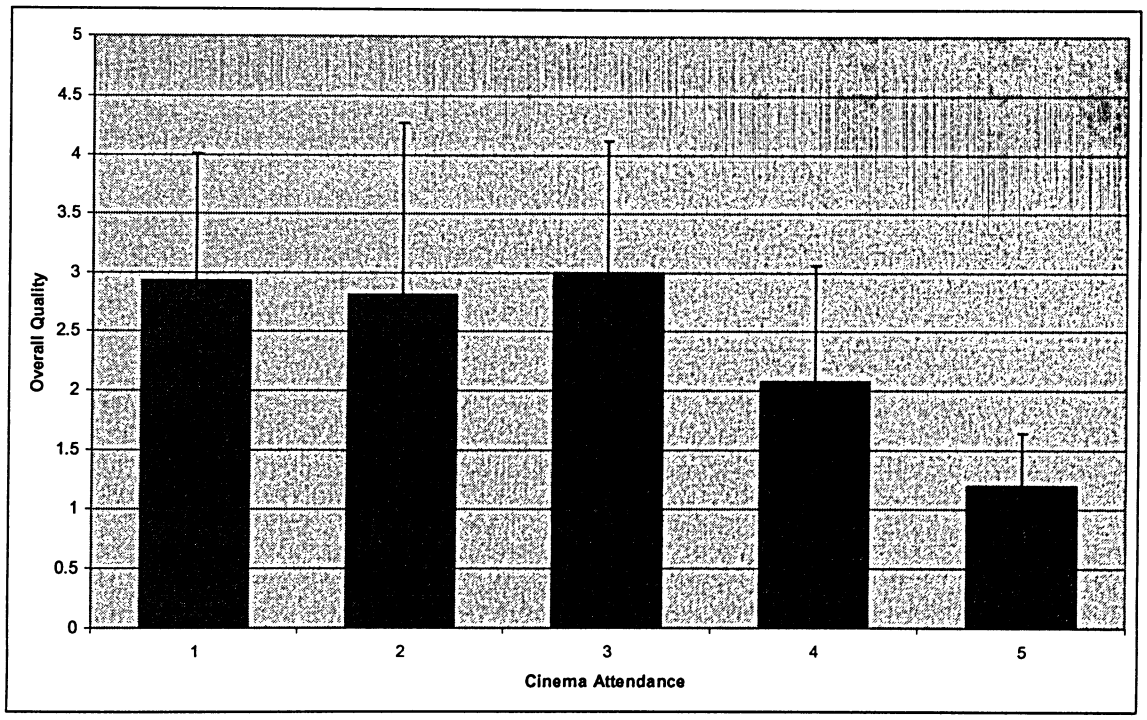


Figure 11 - Mean values and standard deviations for overall quality as grouped by reviewer cinema attendance level

A one way ANOVA showed that there was a significant difference between viewers of differing employment status on the variables of Overall Quality, Listen to More and Compared to Pro (see Table 13). The means values for these groups indicate those with more free time (e.g., those who are unemployed) were more willing to listen to more description while those with less free time where less willing.

Table 14 - Significant differences reported between reviewer employment for overall quality, compared to pro, and listen to more

		df	F	Sig.
overallquality	Between Groups	5	4.089	**.001**
	Within Groups	280		
	Total	285		
comparedtopro	Between Groups	5	4.914	**.000**
	Within Groups	268		
	Total	273		
styleofdelivery	Between Groups	5	2.181	.057
	Within Groups	276		
	Total	281		
listentomore	Between Groups	5	2.944	.013
	Within Groups	279		
	Total	284		

As seen in figure 12, University and College students showed a higher mean value ($M = 3.14$, $SD = 1.24$), for the rating of Listen to More (rated on a 5-point Likert scale where 1 was XX and 5 was YY), people who were unemployed had a mean rating of Listen to More of ($M = 2.95$, $SD = 1.30$) and Retired ($M = 3.35$, $SD = 1.32$) the mean was the lowest for people who Work at Home ($M = 2.48$, $SD = 1.45$) and reviewers. Moreover, a similar trend can be seen with Overall Quality.

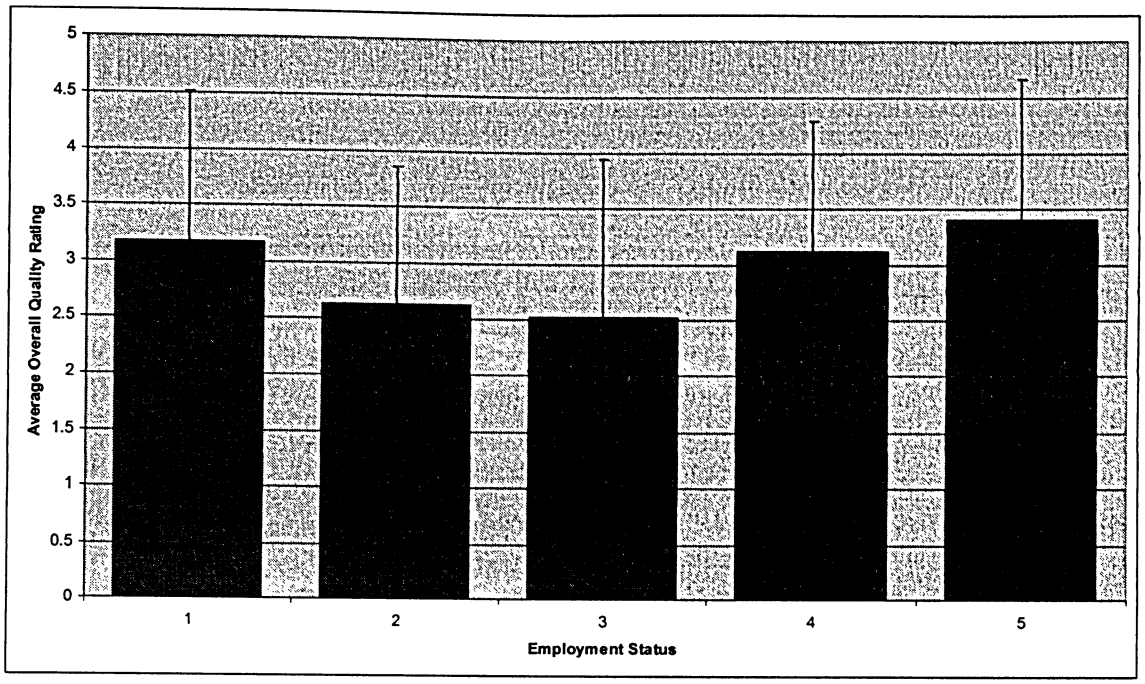


Figure 12 - Mean values for over quality grouped by reviewer employment status

An ANOVA was carried out to examine differences in opinions between genders. There was a significant difference in the opinion of vocabulary between male and female viewers [F(1,279)=5.221, p=.023]. Females rated the level of vocabulary ($M = 3.07$, $SD = 1.17$) as lower than their male counterparts ($M = 3.38$, $SD = 1.06$). There were no other significant differences between male and female respondents.

There were no significant differences in reviewer opinion for Age, Movie Access Strategy, Familiarity with *The Daily Show* and Opinion of *The Daily Show.*

5.3 Phase I and Phase II Data

The overall quality groupings found from the data analysis of phase 2 data was used to discover differences between of good, medium and bad describers. There was significant difference of *text and space* usage found between the poor group of describers

and the medium and good describers [$F(2,7) = 8.357, p = .014$]. This indicates that either the use of the text and space tools promotes the creation of good quality descriptions, or good describers are naturally attracted to the use of these tools.

A Mann-Whitney test was conducted to determine the effect of description length and number of descriptions on describer ratings. However the results were not significant. This is likely due to the low number of phase I participants. Further study is needed to explore the effects of the individual functions of LiveDescribe on description quality and describer preferences.

6. Discussion

6.1 Phase 1

Several key discoveries emerge from the analysis of the pre and post study questionnaire data of phase I regarding the usability of LiveDescribe. Learning to use the LiveDescribe software was reported as being an easy task (10 of 12 participants (83%) reported Learning to Use LiveDescribe as easy or very easy). In addition, using the computer itself for description was also rated as an easy or very easy task by all describers. These results suggest that as a tool, LiveDescribe is easy to learn and that using a computer-based system will likely have potential as a useful tool for creating video description by amateur describers.

All describers in Phase I of the study were also able to complete the task of description even though most of them had little or no exposure to video description prior to their participation. This indicates that is possible for amateur describers to carry out descriptions tasks and that tools such as LiveDescribe might be helpful in facilitating the completion of that task. It appears that the implementation of the LiveDescribe interface then meets the first principle of universal design, equitable use and the third principle, simple and intuitive use since all participants who possessed a wide range of skills and experience reported LiveDescribe to be easy and useful. Since all participants were new to both description and LiveDescribe more research is needed to determine the usability level of LiveDescribe for advanced users.

The high mean values indicating a positive response to the usefulness of the Graphical Timeline, Automatic Space Detection and Timer for Recording are also strong

support for the usability of these LiveDescribe interface functions. Ten of 12 (83%) participants reported these as useful or very useful for their creation and recording of video descriptions. This result provides some evidence that these features assist in making LiveDescribe usable and useful for the task of creating description. The fourth principle of universal design, perceptible information, appears to have been somewhat successfully implemented into LiveDescribe as the primary information relying functionality of LiveDescribe, the graphical timeline, was reported as useful. LiveDescribe delivers its principal information through a graphical mode as seen in the timeline as well as in a text mode and more research is needed to explore the benefit of delivery this information in two different modes.

These positive results suggest that we should continue to develop LiveDescribe as it appears to be meeting some of the criteria of universal design and usability. First time users and, more importantly, first time describers found it to be easy to use and learn, and useful. However, several usability issues where discovered and several system bugs uncovered. This suggests that the fifth principle of universal design of tolerance for error has not yet been achieved, however, all of the problems identified can be resolved. All known bugs experienced during the Phase I trials have been addressed but further efforts must be made to handle user generated or conceptual errors using improved error handling techniques such as avoiding them altogether or using better error and recovery messages.

One common usability issue that arose during many phase I trials, was participants forgetting to save text entered before they recorded their description or move

to another space. A reminder message or an automatic save feature would prevent this error.

The data collected from the thematic video analysis provides some insight into how each respective describer approached and completed the task of description. Some describers preferred the use of the text and timed recording tools and used it extensively, while others chose a more manual approach to description and ignored those tools just as extensively. Almost all describers used either automatically detected or manually defined spaces when creating their description suggesting that this feature is perceived as useful and is used in practice. These observations suggest that the second principle universal design, flexible use, has been successfully implemented into LiveDescribe. Users could choose from among three basic description creation strategies; a completely manual approach, a combined manual and automatic approach, or a fully automatic approach. This flexibility supports novice users who may want to take advantage of the automated tools of LiveDescribe as they are learning as well as advanced users who may choose a more manual approach.

When combining the thematic data from Phase I with the opinion of viewer data from Phase II some of the most interesting and useful patterns emerge. It seems that either good describers are attracted to the use of the text and spacing tools, or that these tools facilitate good description techniques. The poorer rated describers seem to choose a manual or free approach and did not use the text or spaces tools to the same degree as the higher rated describers. Unfortunately, due to technical problems the phase I thematic analysis data analysis is lacking data from describer D12 and describer D6. Having

complete data sets for these two describers may have provided further evidence t the results found. Further study is needed to elaborate on these preliminary findings.

The final two principles of universal design, low physical effort, and size and space for approach, were not evaluated with users in this thesis as no describers with physical limitations were available.

6.2 Phase II

6.2.1 Describer Groupings

Several surprising findings arise from the analysis of phase 2 data. First, it seems that within the study, describers could be categorized as "good", "medium" and "weak" describers based on the overall quality ratings where a good describer is rated above an overall quality rating of 3.37 out of five; medium is rated at an overall quality rating between 2.5 and 2.36 out of 5 and poor is 2.5 out of five or below. As seen in Figure 3 describers D12, D10 and D9 fit into the good category. Describers D11, D8 and D4 appear to be the least preferred describers and their ratings for Overall Quality tend to fall into the poor category. The remaining describers fall into the medium category and lie somewhere in between the results of the "good" and "bad" describers. Essentially, this finding begins to answer one of my research questions regarding the feasibility of the process of description by amateur describers. Indeed, at least 3 and possibly more of the 12 describers used in phase 2 were able to create descriptions that the majority of the audience members rated as having good quality. One important next step in discovering how amateurs describers make good quality description is to examine the common

characteristics that arise in the process. Although, this research did not specifically set out to identify these characteristics, there is some indication of them in the data collected.

6.2.2 What Makes a Good Describer

At first glance describers D12, D10 and D9 appear to have a limited number of description related characteristics in common. Describers D12 and D10 are female, while describer D9 is male. Describer D12 is in the 30-39 age category while describers D10 and D9 are in the 19-29 years age category. Describers D10 and D9 were either familiar or very familiar with *The Daily Show* and liked it very much, while describer D12 had never seen the show before. None of three top describers was very highly educated, with describer D12 holding a college diploma, and D10 and D9 holding high school diplomas. While describers D9 and D10 watch over 15 hours of television a week, describer D12 watched none. This may suggest that describer age, education, television viewing habits and familiarity with the program are not important factors influencing the quality of description produced although further study with more participants and a more targeted experimental design regarding these specific parameters is required.

It was surprising that familiarity with the program contents was not a factor that appeared to have an impact on describer performance or quality ratings. I would suggest that further studies that examine the relationship between describer familiarity with the show and viewer quality ratings for different genres are required. It may be the case that the relatively simple nature of *The Daily Show* allowed describers who were not familiar with the program to create good quality description regardless of their opinion of the

show, or that it is easier to produce high quality descriptions for comedy. *The Daily Show* is highly dialog based, with only a few visual elements that are crucial to the delivery of the program. It may be that lower quantities of description and lower levels of complexity in the description are required for this type of program for viewers to understand the show. Complex programs with multiple characters and settings, sub-plots or intricate actions scenes may require more familiarity to produce high quality descriptions. Further study using different genres of programming is required to elaborate on these findings.

While there was little demographic similarity between the three top-rated describers there were important similarities in the description type, style and number. All three top-rated describers had a similar number of descriptions (35 for D9, 25 for D10, and 26 for D12) and total description lengths (122.79 s., 113.96 s., 165.73 s for D9, D10 and D12 respectively). This suggests that these describers found the length, number and positioning of descriptions most appreciated by the audience. They very likely inserted a description when it was required and did not insert a description into places it was not required, while being careful not to overlap their descriptions with important dialog.

Describers 9 and 10 used the space and text tools extensively to create descriptions. D9 used the space and text tools almost exclusively while describer D10 used spaces for nearly all descriptions. Video analysis data was not available for D12 so her use of space and text tools was not available for comparison, however from my observational notes, she did use the text features quite often.

The results from D9 and D10, however, suggest that good describers used the text and space tools, where the weaker describers did not. What is not obvious from the data is whether the people who are inherently good describers would be successful regardless

of the describing system used, or whether the use of a computer-based tool enables good description. Further studies which would examine various automatic and manual tools between describers already found to be able to produce high quality description are required to determine the value of LiveDescribe.

All three describers seemed to enjoy the process of description as they rated their experience of creating description as good or excellent. For example, Describer 12 stated that "The task was not a difficult one if anything I enjoyed it a lot."

There were also some stylistic similarities between the high-rated describers. Describer 12's style followed very closely to that of the traditional video describer in that she used a third person narrative form and her enunciation was very clear. However, her tone was very bright and cheerful, and it seemed to match the overall tone of the program similar to the way sport commentary appears. Her descriptions where thorough and very detailed, and rarely interfered with the existing dialog. They also seemed to have sufficient detail and the nuance matched what a sighted audience might experience. For example, when Jon Stewart does an impression of George Bush, she does not describe it is as "Jon Stewart imitates George Bush", instead she says "Jon slouches over with hands out-stretched in an unsure fashion while squinting his eyes." This more subtle delivery allows for the audience to come to their own conclusion that Jon is indeed imitating George Bush, which I believe is a critical part to the humor. Describer D10, also a female, maintained a minimalist style, rarely describing over dialog and describing only the minimum amount required. For example, during one scene of *The Daily Show* an image of a tiger wearing a pimp's costume is described by Describer D10 as "The Exxon Tiger has some 'bling-bling'", rather than trying to describe all the elements of the

costume including jewelry, furry hat and sunglass. Using the term "bling-bling", a pop culture reference to flashy jewelry, stays more within the light natured feel of the program rather than trying to coldly describe the scene objectively. In addition, by using a pop culture reference, the description is delivered in a more compact form. D10 also had a soft sounding, non-intrusive tone of voice and it may have been a factor contributing to her high quality score. More study is needed to explore the relationship between quality perception and voice intonation.

D10 also works as a live captioner at universities and colleges, and this influence may have given her greater insight into the needs of the blind and vision impaired audience members as it is most like part of her duties as a cautioner to understand the requirements of accessibility. Further research is required to study to influence of careers in the disability field. Describer D9, a male, also maintained a minimalist style, and like D10 described in a softer, non-aggressive, non-intrusive style of voice. D9 also rarely spoke over dialog and for the most part was able to fit description with the spaces without dialog available. Like D10, D9 also used pop-culture references to help shorten his descriptions, for example he referred to "Puff-Daddy" when describing the Exxon Tiger's pimp costume.

6.2.3 What Makes a Poor Describer?

Describers D4, D8 and D11 all seem to fit in the poor describer category as their mean overall quality ratings were below 2.3. One interesting common characteristic among these describers is that they all spoke with a noticeable accent. Describers D4 and D11 were ESL speakers, D4 being a French speaker while; D11 spoke with a thick East

Asian accent. Describer D8, while a native English speaker, spoke with an Australian accent. Describer D11 even predicted that audiences would not appreciate her accent and when asked how an audience would like her description she commented "I think they will not even understand my heavy accent to it." This result, although not conclusive, suggests that description audiences may prefer describers who speak without an accent in the language used for description, or more specifically they would like to hear a describer speaking in the same vernacular as their own. Due to the relatively low number of describers with a discernable accent, it is difficult to determine whether this is the most important factor contributing to the poor quality ratings of this group and more study must be conducted in order to determine which factors most contribute to a poor audience rating. I would suggest however, that describer accents which interfere with the comprehension of the description will likely interfere with audience enjoyment levels.

Other factors which seem to contribute to levels of perceived quality are description length and total number of descriptions. Describers D8 and D4 have lower than average total description lengths (53.04 s, 34.72 s), and total numbers of descriptions (18, 12) while describer D11 have the highest description length (468.80 s) and total number of descriptions (60). This result suggests that extreme description lengths, either too short, or too long, and total number of descriptions, too few or too many may result in descriptions that are problematic for audiences. Describers D8 and D4 likely missed many possible description opportunities while describer D11 likely described over dialog or described when it was not necessary, causing the low overall quality rating. This is further supported by the fact that describer D11 used available spaces only with 15 of 58

of her total descriptions, while the remaining descriptions were placed over areas where the system reported dialog.

Describers D11 and D4 were the only describers to not use the text tool at all, which suggests that either use of the text and space tools have a positive effect on the overall quality ratings or that good describers tend to want to use them and poor describers tend to not want to use them. Describer D8, however, did use the text tool for nearly all of her descriptions, which may then suggest that there are other factors such as length of description, timing and understandability of the describer that may have a larger impact on ability of describers to produce high quality descriptions.

Error rates made by describers did not seem to affect the over quality level of the final description. The poorest describers have the lowest errors rates (D1 = 0, D4=2, D11 = 5) and medium or good describer have higher rates (D3 = 25, D9=16). It appears more likely that error rates are tied to the total number of descriptions or total time spent describing rather than overall quality, as with an increase in the number of actions there is likely an equally portioned increase in error rates.

6.2.4 Differences Between Reviewers

One surprising result from the phase II data is that describer ratings were related to vision status of the reviewer. Sighted participants seemed to rate description quality as lowest in general, followed by raters with low vision and finally blind reviewers. This result may be due to the lack of available described material for blind audiences and that

any described content was seen as a positive step. As a result, they may have been more likely to rate description quality as high.

Low vision and sighted viewers, however, had lower ratings likely because they can see, at least to some degree, what the description is or is not missing. They would therefore be more likely to rate descriptions as poorer when they are mistimed, have missing words or when descriptive elements are otherwise incorrect.

The significant difference discovered between reviewers grouped by how often they attend the cinema was expected. I suggest that reviewers who attended the cinema more often, rated description more harshly because of an increased exposure to expectations of programming and video description. Increased exposure to high quality professional video description most likely leads to increased expectations of description quality in general. Those who rarely attended the cinema and therefore were rarely or never exposed to quality video description were likely satisfied with any form of description with which they are presented.

6.3 Video Description Process

One of the major research questions posed by this thesis related to the process of video description creation mediated by LiveDescribe as a viable and advantageous work flow process. Several discoveries arose from the findings of this thesis. First, every describer included in Phase I data was able to complete the task of describing an entire twenty minute episode of the Daily Show in one session. The describer which took the most time was describer D9 who completed the task in just under 2.5 hours, while some

describers were able to finish within one hour. This means a time requirement of 3 - 7.5 hours per hour of programming is needed for amateurs with no training, and no experience in description to create description. This amount is much lower than the requirements reported in the literature of one work week to describer 1.5 to 2.5 hours of content. This is a very promising result due to the fact that each describer had little, if any, description experience and no experience with LiveDescribe. It seems that novices are able to complete the difficult task of description in a reasonable amount of time. In addition, most people were able to complete their description task using LiveDescribe with few errors, the most common being text saving and navigation error.

As users become more practiced and experienced with the description process, it is expected that they will become more proficient users of LiveDescribe. It is also expected that describers will make less use of the text entry process as they become more practiced with the description process. I feel that as novice describers gain experience they will become more familiar with the requirements of their description and will likely be able to create description without pre-writing them first. The text entry functionality may therefore scaffold the learning process for new describers in a way that allows them to become more proficient in a short period of time. Future study is recommended to determine the impact of various LiveDescribe functions on their contribution to learning and becoming proficient at describing.

The majority of the describers used the automatic space detection tool and a majority of them reported it as being useful or very useful. This suggests that having automated space detection is a desirable function for description. Similar to the text entry tool, automated space detection may also facilitate the learning process for describers.

Automatic internet publishing capabilities which will soon be incorporated into LiveDescribe will streamline the process even further and allow describers to publish their descriptions to a wide audience, something which is currently not readily available.

The results discussed here suggest that a video description wiki facilitated through LiveDescribe may be feasible as one of the foundational principles assumptions in developing such a technology requires that amateur describers are able and willing to create descriptions that could be appreciated by blind audiences. In this thesis, first steps were taken to show that amateurs can create description that blind audiences would consume that description. These amateur describers were able to create description with little training and within a relatively short timeframe compared with the time required for professional description to be created. Combined with the fact that friends and families of blind viewers are motivated to provide informal description without monetary compensation as reported by Schmeidler (2001), it seems feasible that a description wiki is possible. If descriptions could be shared using a wiki-based environment, more blind viewers could benefit from the informal descriptions created by any amateur describer. They could also have access to different versions of description because various amateur describers could contribute different versions of the same show. This would then allow viewers choices and opportunities to have preferences (helping to meet the second principle of universal design).

7. Limitations

While this study indicated some interesting and promising findings, there are important. One important limitation is that this study only examines description created for a single 20 minute show, *The Daily Show*, from one genre, a comedy sketch/talk show. Further study must be conducted to examine whether description with adequate quality can be created for different shows in the same genre as well as shows, movies, music videos or other video from other genres and the impact that those descriptions and program types have on audiences and their perception of description quality. It is possible that had this study been conducted using a more visually complex show requiring more description or more complex description, the results would have been much different. I chose *The Daily Show* specifically because it seemed to have enough between-dialog spaces for description opportunities and it was short enough that descriptions for the entire show could be created in a single one-hour session. It is possible that had the program required more description, participants would have become bored or fatigued and not been able to complete their description tasks in a single session. Further research is required to: 1) understand the impact of different genres and description complexity on the ability of amateur describers to create description; and 2) whether changes made as a result of this study improved the usability of LiveDescribe.

Another limitation of this study is that only one episode of one 20-minute show was used. The novelty of the concept and/or the show may have influenced the interest and motivation of the describers to complete their descriptions tasks and audience's willingness to give positive or negative quality ratings. It is possible that ratings for description could have been much different had reviewers judged several episodes of *The*

Daily Show, or multiple episodes of longer shows. Description characteristics which may be desirable for this single show, may not be desired once the novelty effect wears off, or once reviewers become more familiar with the show. I would recommend that further research be conducted to study the novelty effect of amateur description over long periods of exposure. In addition, reviewers were exposed to five different describers during one viewing of *The Daily Show.* Had each reviewer been presented with the version from only one describer, the quality rating may have changed. To examine this question, I would recommend a much larger longitudinal study involving many different program genres and describers.

In order to allow participants to complete an entire episode of description for a single show, a program with relatively low quantity and complexity of description requirements was chosen. Had a longer, or a more visually intense and less dialog driven program been used, participants may have not been able to complete the description task in a single session. In addition, audiences may have become frustrated or bored with a long show that had poor description and this could have negatively affected their willingness to participate in this initial exploratory study.

To explore these two limitations a wide spread internet based experiment could be implemented. Such an experiment could be facilitated by LiveDescribe and a publicly available description wiki. Amateur describers could download LiveDescribe, use their own copy of a movie or television program, create descriptions for that program and then publish only the description to the description wiki. Blind, low vision and sighted users could then visit the description wiki, download the descriptions and synchronize them with their own version of the movie or television program. Questionnaires or a rating

system could be integrated into the site and would help filter the better quality description from the lower quality as well as serve as a valuable research tool. The description wiki could easily be leveraged to collect large amounts of data regarding the perceived quality levels of the description or any other aspect of description that warrant study.

The Daily Show is a dialog based talk show, meaning that the show contained much dialog and, according to the detection algorithms in LiveDescribe, 186 seconds of program were opportunities for description. *The Daily Show* s 21 minutes long or 1260 seconds long and about 14% percent of the show contained spaces for description opportunities. This only represents the possible points for description insertion, however, and does not give a measure of the visual complexity of the program.

For this initial exploratory study, I did not want to over exert participants but I still wanted to ensure that there were multiple opportunities for a variety of (visual content, length and complexity) description. Had describers been required to create description for a program that requires more description or more complex description, it may have caused high levels of fatigue in participants and this could have had a negative impact on the quality of their descriptions. Future studies may be designed to examine the fatigue rates of and the resulting impact on description quality for amateur describers by varying the visual complexity of the show and the number and length of description opportunities. Fatigue could be measured by user reports, heart rate analysis or other physiological measurements.

Another limitation of this study is the low number of participants in phase I. Although more participants were not necessarily needed to answer the main research questions presented in this thesis and measures, , any results from the phase I study must

be considered conservatively. However, it is common in usability methodology for small numbers of participants to be used since the purpose of usability testing is to identify user problems or issues with interfaces (Nielsen, 1995). In addition I did not use phase I data as my primary source for analysis. Future study with more describers as mentioned in section 6 is needed to help determine specific usability issues surrounding LiveDescribe and also to determine what describer and description characteristics audiences appreciate.

Another major limitation of the study is the trustworthiness of the data obtained from online sources. Although users labeled themselves as blind or low vision I have no way to confirm these claims. It is possible, however extremely unlikely, that only a few people answered the survey many times, making the data unsound. However, I found that the reviewer comments differed in many ways such as level of candor, writing style, and number of issues addressed that and I was therefore convinced that the responses were genuine.

In addition, many of the responses were candid and harshly critical of the description and the program content, to a level of which I have never seen, and would not expect in an experiment where the researcher and participant are face-to-face. I therefore suggest that the online data was actually more accurate due to a level of honesty that would be unexpected in a face-to-face study due to social manners and expectations in a face-to-face meeting. Further study is needed to compare the results of online description reviews and face-to-face reviews to determine the effect that a face-to-face experiment has on quality rating.

A final limitation was that a prototype of LiveDescribe containing undiscovered errors was used for phase I. During the phase I study, new coding errors were discovered.

Although every attempt was made to correct the errors as they appeared, most participants experienced at least one code bug. One code error that was particularly problematic prevented users from recording over areas where a previous description had existed. Participants were instructed to save their work often and if major error messages, erratic system behavior, or a system crash occurred they were to reopen their project and continue as normal. Although I believe this solution helped to alleviate the effects of these system errors on the final descriptions produced by each participant, there was a negative impact from these bugs on users. I observed users becoming frustrated with certain features of LiveDescribe. For example the text recording tool functioned inconsistently and as a result participants may have changed their work pattern to accommodate the error, in a way which they may not have, had those errors not been present. However, since these errors affected text saving and recording tools, and users still attempted to use them and reported them as at least moderately useful and easy to use, I suggest that these tools are indeed helpful. Once these errors are addressed, usage levels and perceived ease of use may increase even more so. I cannot, however, make definitive conclusions on the description process using LiveDescribe from the data of this research because of the impact on the description creation process caused by these bugs. A new round of user testing with a more stable version of LiveDescribe is recommended.

+++

8. Conclusion and Recommendations

The results from the studies discussed in this thesis provide some answers to the research questions posed in Chapter 3 and are discussed in this chapter.

Research question one relates to whether amateur describers are able to create video description with acceptable quality levels. The first phase of the study showed that amateur describers can create descriptions for blind and low vision audiences with little or no training in video description techniques. Phase II of the study then showed that there were definite preferences for certain describers and their descriptions. Describers D12, D10, and D9 were identified by blind, low vision and sighted audiences as having the best overall quality. These preferences seem to be based on the describer's vernacular, the describer's tone of voice, and the length and timing of the description.

In addition, blind, low vision and sighted audiences also showed a definite dislike for specific describers and their descriptions. Describers D11, D8, D4 had the lowest quality ratings among the blind and low vision audience in this study. There were very few common or obvious factors, other than language accent, that stood out as the cause of the low ratings. There are some factors such as over or under describing and non-use of the text or space tools provided by LiveDescribe that appeared in two of the three describers.

Further research must be conducted in order to determine the exact effects describer aspects have on the perception of quality and definitely identify the various factors that contribute to description quality and user preferences. Description or describer factors such as description style, emotional contents, linguistic complexity, flow, cadence, volume, overlap of dialogue and quantity of description all likely

contribute to audience preferences. Audience factors that could have an impact on quality or preference ratings could include literacy levels, gender, and tolerance for overlap of dialog, genre preferences and television or movie viewing habits.

The second research question relates to whether the process of creating description, mediated through LiveDescribe, is viable and practical for amateur describers. In phase I of the study, it was found that the majority of the describers were able to complete the challenging description task given to them, and reported that the tools in LiveDescribe that they employed to assist in this task were easy to use and useful. Phase II data showed that at least some of the describers were able to not only produce description but that their descriptions seem to be perceived as having an acceptable level of quality. I suggest that had the describers in phase I not been given the LiveDescribe software to assist them in creating description that the majority of them would have found the manual process required for description to be tedious and too difficult to complete as they lacked the necessary technically skills to manually find, mark and count possible description spaces as well as produce the description text to fit in those spaces and insert it between the appropriate time code markers.

To further improve the description creation process, and to assist amateur describers in preparing description, training materials and guidelines could be prepared for distribution with software such as LiveDescribe. Recommendations for the process of description could include basic summary research information, such as user preferences for style of delivery, language use and timing. In addition, tutorials on the usage of text and timing tools of LiveDescribe should be provided alongside the benefits of using such tools.

I believe, however, the most valuable resources will not be a set of guidelines but rather other amateur or professional describers that would be present on a wiki. As a publicly accessible description wiki would be designed to filter the most popular description to the top, other describers would be able to view that description and understand what blind audiences appreciate in a describer. A description wiki could contain many different types and styles of descriptions and audiences could have choices regarding which description they consume. A rating or user generated recommendation or comment system could be employed to offer user opinions of the various description choices so that new users could make informed choices about the descriptions they consume. Describers could observe what other describers do and what audiences are reporting, and incorporate this information into their own styles. As video description, unlike closed captioning in its current form, is a subtle art form, much like writing, music composition or film making, guidelines may not be the most appropriate method of transferring description skill sets from the experienced describer to the novice. Guidelines are likely too restrictive and therefore could not possibly provide accurate and appropriate recommendations for each and every possible description situation.

The final research question in this thesis was, also addressed in phase I and confirmed in phase II. Describers reported LiveDescribe was useful and easy to use. This reported opinion was confirmed through video analysis which did indeed confirm that the core functionality of LiveDescribe, the automatic space detection, text entering and saving functionality and voice recording features were used by the describers, specifically the describers who produced the highest quality description.

The viability and usability of LiveDescribe as tool for assisting in the creation of video description has also been demonstrated in this thesis, however, there were a number of usability issues that were identified with the software. These include video navigation and text entering and saving sequencing. Some users experienced difficulty using the navigation bar and others reported issues with navigating between pages of the timeline. Other users had problems saving text before recording which resulted in lost text. The majority of usability and coding problems are simple to rectify. For example, a message box alerting the user that text has not been saved, or automatically saving the text would be simple to implement and would alleviate this relatively major interface problem.

Improvements in LiveDescribe include discrimination algorithm development, increased support for additional video formats, including DVD, Blue-Ray and HD-DVD as well as single click publishing capabilities that will upload description projects direction to the LiveDescribe description wiki.

The results of the studies and development work carried out in this thesis, I suggest that amateur describers can produce description at a quality that blind viewers can appreciate. I also recommend proceeding with development of a video description wiki, facilitated by LiveDescribe, which would be supported primarily by amateur created video description.

References

Barribeau, P., Butler, B., Corney, J., Doney, M., Gault, J., Gordon, J., et al. (2005). Survey research. Writing@CSU. Colorado State University Department of English. Retrieved April 30, 2007 from *http://writing.colostate.edu/guides/research/survey/.*

Bleicher, P. (2006). Web 2.0 revolution: Power to the people. Retrieved April 10, 2007 from http://www.actmagazine.com/appliedclinicaltrials/article/articleDetail.jsp?id=363637&searchString=web%202.0.

Branje, C., Marshall, S., Tyndall, A., & Fels, D. I. (2006). LiveDescribe. Americas Conference on Information Systems. 4-6 August 2006 (pp. 3033-3041). Acapulco, Mexico.

Clark, J. (2007) Media Access. Retrieved April 16, 2007 from http://joeclark.org /access/crtc/crtc-ad.html.

Canadian Radio and Television Commission. (2004). Broadcasting Public Notice CRTC 2004-2. Ottawa: Canada.

Dahan, E., Srinivasan, V. (1998).The predictive power of Internet-based product concept testing using visual depiction and animation. *Journal of Product Innovation Management, 15*(2) 99-109.

Doyle, J. K. (2001) Introduction to survey methodology and design. Retrieved from January 15, 2006 from http://www.sysurvey.com/tips/introduction_to_survey.htm.

Fels, D. I., Udo, J. P., Diamond, J. E., & Diamond, J. (2006). A comparison of alternative narrative approaches to video description for animated comedy. *Journal of Visual Impairment and Blindness, 100*(5), 295-305.

Hammontree, M.(1994) Remote usability testing. *Interactions, 1*(3), 21-25.

International Telecommunications Union. (1998). Interactive Subjective audiovisual quality assessment methods for multimedia application. (ITU-T P.911). Geneva: Switzerland.

Long, S. (2006). Exploring the wiki world: the new face of collaboration. *New Library World 107*(3/4), 157-159.

NCAM. (2003). MAGpie 2 Documentation. Retrieved on December 27, 2007, from http://ncam.wgbh.org/webaccess/magpie/magpie_help/.

Nielson, J. (1995). Advances in Human-Computer Interaction. United Kingdom: Intellect Books.

Peli, E. (1996). Evaluating visual information provided by audio description. *Journal of Visual Impairment & Blindness, 90*(5), 378-85.

Pettitt, B., Sharpe, K., & Cooper, S. (1996). AUDETEL: Enhancing television for visually impaired people. *British Journal of Visual Impairment, 14*(2), 48.

Pfanstiehl, M., & Pfansteihl, C. (1985). The play's the thing: Audio description in the theatre. *British Journal of Visual Impairment, 3*(3), 91.

Piety, P. J. (2004). The language system of audio description: An investigation as a discursive process. *Journal of Visual Impairment and Blindness, 98*(8), 453-469.

Pinquier, J., Rouas, J., André-Obrecht, R. (2002). Robust speech / music classification in audio documents. ICSLP-2002, 2005-2008.

Preece, J., Rogers, Y., Sharp, H., Benyon, D. (1994): Human-Computer Interaction. United Kingdom: Addison-Wesley

Ramos, M., & Piper, P. S. (2006). Letting the grass grow: Grassroots information on blogs and wikis. *Reference Services Review, 34*(4), 570-574.

Schmeidler, E., & Kirchner, C. (2001). Adding audio description: Does it make a difference? *Journal of Visual Impairment & Blindness, 95*(4), 197-212.

Sierkowski, B. (2002). Achieving web accessibility. Proceedings of the 30th annual ACM SIGUCCS Conference on User services, November 20-23. 2002 (pp 288-291). Providence, RI, USA.

Softel-USA. (2001). ADePT Audio Description Technology. Retrieved December 22, 2007, from http://www.softelusa.com/l4_products_subtitling_adept_main.php3.

Statistics Canada (2001). Participation and Activity Limitation Survey. Retrieved Dec 27, 2007, from http://www.statcan.ca/Daily/English/021203/d021203a.htm

Story, M.F. (1998), Maximizing usability: The Principles of Universal Design. *Assistive Technology Journal, 10*, 4-12.

Stvilia, B., Twidale, M., Gasser, L., & Smith, L. (2005). Information quality discussions in wikipedia. Proceedings from ICKM '05: The 2005 International Conference on Knowledge Management. Kuala Lumpur, Malaysia.

The Independent Television Commission. (2000). ITC guidance on standards for audio description. Retrieved January 16, 2008 from http://www.ofcom.org.uk/static /archive/itc/itc_publications/codes_guidance/audio_description/index.asp.html

Treviranus, J., Silverman, C., Fels, D., Boyce, M., Rosen, M., Diamond, Sara., et al. (2004). Online Enhanced Captioning Guidelines. Retrieved Dec 27, 2007, from http://cnice.utoronto.ca/guidelines/caption.pdf.

Appendix A: Information and Consent Form

Project Title:	LiveDescribe – Can Amateurs Create Quality Video Description?
Principal Investigator:	Carmen Branje, Graduate Student, Ryerson, Masters of Management Science yerson.ca
Supervisor:	Deborah Fels, P.Eng., Ph.D., Ryerson University (416)-979-5000 ext. 7619 or dfels@ryerson.ca

Consent to Participate in Study from Subject

Information Form

In this study, we would like to determine whether amateur video describers can create description for blind or low vision audiences. To do this, you will use special software, called LiveDescribe, designed to assist people in creating descriptions for video content.. If you agree to participate in the study, you will be trained on how to use LiveDescribe and then asked to view 20 minutes of television content and create video description for that content. You will also be asked to complete two questionnaires; one prior to beginning the study and one after. These questionnaires will ask for your impressions and opinions of video description in general, and on your experience of using LiveDescribe to create video description.

Confidentiality
All data will be confidential and will not be published except as summary data. All data will be used exclusively for research and educational purposes. Only the principle investigator, Carmen Branje, and his supervisor, Dr Deb Fels, will have access to this data. Anonymous data not associated to yourself in anyway will be made public through the release of the findings of this study. Videotapes and other data will be kept in a locked in a location at Ryerson University for five years and then destroyed.

Risks and Discomforts
The risks associated with the study are very minimal. You may experience some fatigue while creating video description. A scheduled break will occur following the training sessions, but you may also take other breaks during the description session if you become fatigued. You may also discontinue the study at any time without penalty. You may also experience some discomfort in answering some of the personal questions in the questionnaire such as age and gender. You are not required to answer any question that makes you uncomfortable. Any data collected will only be used to analyze the feasibility of using LiveDescribe to create video description. Only members of the project team will have access to all the data, which will be stored in the project office in a locked filing cabinet.

Expected Benefits
Though subjects will not directly benefit from their participation in this study, the data will be used to assess the limitations of the current method of creating and distributing video description and likely result in an increase in the available of video description for blind and low vision audiences. Final copies of findings of the study can be requested from Carmen Branje (cbranje@gmail.com) once such findings are complete. Subjects wishing to receive a copy of this report should leave his or her email address on this consent form.

Costs and/or Compensation for Participation: Travelling costs to and from the study location will be compensated to a maximum of $10.00. Costs will be compensated even if subjects do no complete the entire experiment. This research is funded through NSERC.

Voluntary Nature of Participation:
Participation in this study is voluntary. Your choice of whether or not to participate will not influence your future relations with Ryerson University. If you decide to participate, you are free to withdraw your consent and to stop your participation at any time without penalty or loss of benefits to which you are allowed. At any particular point in the study or doing the performance, you may refuse to answer any particular question or stop participation altogether. Once data is collected and subjected to analysis it will no longer be reasonable to withdraw your data. If you wish to have your data withdrawn from this study please request so before leaving the experiment. Once you leave the experiment your anonymous data could be included in final analysis.

Questions about the Study:

We sincerely appreciate your co-operation. If you have any questions or concerns, please do not hesitate to call Carmen Branje at or Deborah Fels at 416.979.5000 ext. 7619. In addition to the principal researcher and his supervisor, The Research Ethics Board may also be contacted should there be any complaints or concerns about the project, c/o Office of the Vice President, Research and Innovation, Ryerson University, 350 Victoria St., Toronto, ON M5B 2K3, Tel: 416-979-5042

Project Title:	LiveDescribe – Can Amateurs Create Quality Video Description?
Principal Investigator:	Carmen Branje, Graduate Student, Ryerson, Masters of Management Science yerson.ca
Supervisor:	Deborah Fels, P.Eng., Ph.D., Ryerson University (416)-979-5000 ext. 7619 or dfels@ryerson.ca

Consent form for Subject's Participation in Study

I acknowledge that the research procedures described above have been explained to me and that any questions that I have asked have been answered to my satisfaction. I have been informed of the alternatives to participation in this study, including my right not to participate and the right to withdraw without compromise. As well, the potential harms and discomforts have been explained to me and I also understand the benefits (if any) of participating in the research study. I know that I may ask now, or in the future, any questions I have about the study or the research procedures.

Your signature below indicates that you have read the information in this agreement and have had a chance to ask any questions you have about the study. Your signature also indicates that you agree to be in the study and have been told that you can change your mind and withdraw your consent to participate at any time. You have been given a copy of this agreement.

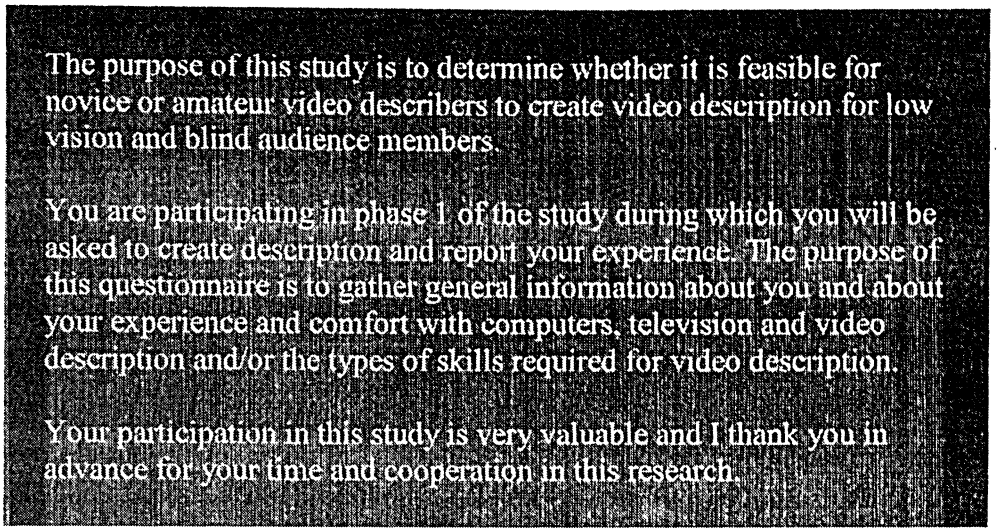
You have been told that by signing this consent agreement you are not giving up any of your legal rights.

_____
Name of Participant (please print)

_____ _____
Signature of Participant Date

_____ _____
Signature of Investigator

Appendix B: Phase 1 Describer Pre-Study Questionnaire



1. What is your gender?

- Male
- Female

2. What is your age?

- 18 to 29
- 30 to 39
- 40 to 49
- 50 to 59
- 60 and over

3. Highest level of education completed?

- ○ High school diploma or below
- ○ College Diploma
- ○ University Degree
- ○ Graduate Degree
- ○ Doctorate Degree

4. Please check any skills you regularly require for your work, hobbies or other day to day activities.

- ☐ Public Speaking
- ☐ Writing
- ☐ Advanced Computer Use
- ☐ Design
- ☐ Organizational skills
- ☐ Interpersonal
- ☐ Technical know-how
- ☐ Domain expertise

5. Rate you level of computer expertise, where an expert is a computer professional such as a programmer or IT manager and a novice is a person who has just begun to use computers.

- ○ Expert
- ○ Intermediate
- ○ Novice
- ○ Never use before

6. How many hours per week do you watch television?

- More than 15 hours of television per week
- 11-15 hours of television per week
- 6-10 hours of television per week
- 1-5 hours of television per week
- I do not watch any television

7. How often do you go to the cinema?

- More than once a week
- Once a week
- Once a month
- Once a year
- Never

8. What is your level of familiarity with the first show, *Odd Job Jack* you are about to describe?

- very familiar
- familiar
- somewhat familiar
- not very familiar
- I have never seen this program

9. You will be working with *Odd Job Jack* in this study. What is your opinion of this show?

- I really like this
- I like this show
- I neither like nor dislike this show
- I dislike this show
- I really dislike this show

10. What is your level of familiarity with the second show, *The Daily Show* you are about

to describe?

- very familiar
- familiar
- somewhat familiar
- not very familiar
- I have never seen this show

11. You will be working with *The Daily Show* in this study. What is your is your opinion of this show?

- I really like this show
- I like this show
- I neither like nor dislike this show
- I dislike this show
- I really dislike this show

---

12. What is your level of familiarity with using video description on television or film?

- very familiar
- familiar
- somewhat familiar
- not very familiar
- I have never heard of description

13. What is your opinion of video description?

- I have a strong liking for video description
- I like video description
- I have no opinion / neutral
- I dislike video description
- I have a strong dislike for video description
- I have never heard of video description

14. What is your experience with creating pre-recorded video description?

- I am a professional describer
- I have created some description, but not regularly
- I have created description once before
- I know how video description is created but have never done it.
- I have never heard of description

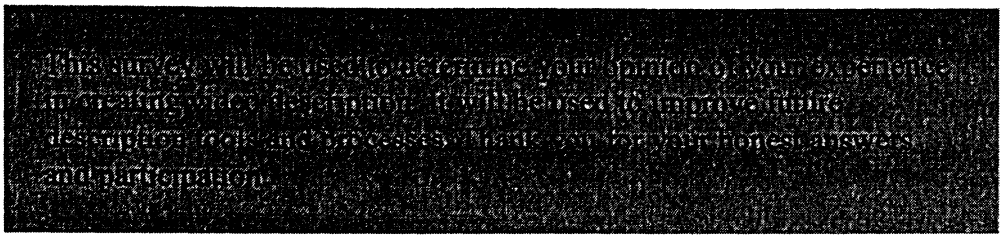
15. Sometimes people will create informal descriptions for blind or low vision friends or family during theatre performances, cinema or television program. What is your experience with creating this type of live description?

- I always provide this type of description
- I regularly provide this type of description
- I sometimes provide this type of description
- I have provided this type of description once before
- I have never provided this type of description before

16. If you have created video description or used it before, please list what you like about it?

17. What are your dislikes of video description?

Appendix C: Phase 1 Describer Post-study Questionnaire



1. Please describe the level of difficulty you experienced in the task of creating descriptions using LiveDescribe for Odd Job Jack.

- very difficult task
- difficult task
- neither easy nor difficult task
- easy task
- very easy task

2. Please describe the level of difficulty you experienced in the task of creating descriptions using LiveDescribe for The Daily Show.

- very difficult task
- difficult task
- neither easy nor difficult task
- easy task
- very easy task

3. Please rate your opinion of the experience of creating video description for Odd Job Jack using LiveDescribe.

- very poor
- poor
- it was ok
- good
- excellent

4. Please rate your opinion of the experience of creating video description for The Daily Show using LiveDescribe.

- ○ very poor
- ○ poor
- ○ it was ok
- ○ good
- ○ excellent

5. How often would you create video description using LiveDescribe again if given the chance?

- ○ often
- ○ sometimes
- ○ once in a while
- ○ never
- ○ don't know

6. How did the descriptions tasks you just completed compare with creating live description during a film, television program or theatre performance for someone with a vision impairment?

- ○ much easier than live description
- ○ somewhat easier
- ○ equal to
- ○ somewhat more difficult
- ○ much more difficult
- ○ N/A

7. Please rate the level of difficulty you experienced with following aspects of creating description?

	very easy	easy	not easy nor difficult	difficult	very difficult
learning to describe	○	○	○	○	○
learning to use LiveDescribe	○	○	○	○	○
Deciding what parts of the show to describe	○	○	○	○	○
Understanding the video material	○	○	○	○	○
Selecting the best words for descriptions	○	○	○	○	○
Using the computer					

8. Please rate the level of difficulty in using following aspects of LiveDescribe to create descriptions.

	very easy	easy	not easy nor difficult	difficult	very difficult
finding spaces to insert description	○	○	○	○	○
editing description boundaries	○	○	○	○	○
creating description boundaries	○	○	○	○	○
writing descriptions	○	○	○	○	○
recording descriptions	○	○	○	○	○
editing descriptions	○	○	○	○	○
navigating through video	○	○	○	○	○
learning to use software	○	◉	○	○	○
understanding graphs					

9. When you were creating descriptions for your show, please rate the usefulness of the following aspects of LiveDescribe. .

	very useful	useful	not useful nor useless	useless	very useless
graphical timeline	○	○	○	○	○
automatic space detection	○	○	○	○	○
ability to write out descriptions	○	○	○	○	○
functions used in recording	○	○	○	○	○
timer for recording	○	○	○	○	○
editing functions for description	○	○	○	○	○
running list of descriptions					

10. How do you believe that people with a vision impairments would react to the descriptions you just created? They would...

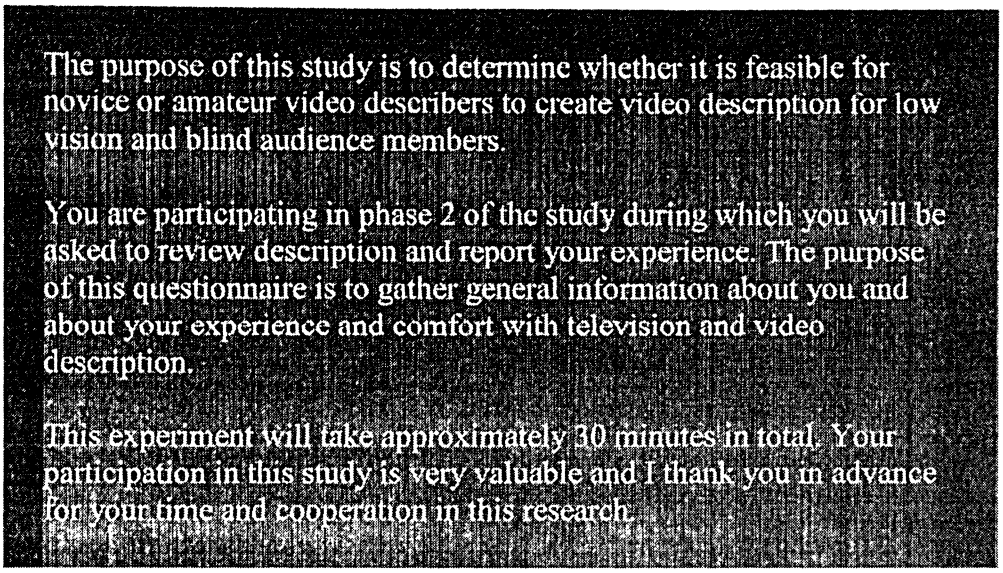
- ○ like it very much
- ○ like it
- ○ neither like it nor dislike it
- ○ dislike it
- ○ dislike it very much

11. Please elaborate on the reasons for your answer to the previous question.

12. Please list the strengths of the LiveDescribe software or the description task

13. Please list the weaknesses of the LiveDescribe software or the description task.

Appendix D: Phase 2 Reviewer Pre-Study Questionnaire



1. What is your gender?

- ○ Male
- ○ Female

2. What is your age?

- ○ 18 to 29
- ○ 30 to 39
- ○ 40 to 49
- ○ 50 to 59
- ○ 60 and over

3. What is your current status

- ○ Highschool Student
- ○ University or College Student
- ○ Work Part-time outside home
- ○ Work full time outside home
- ○ Work at home (includes child care)
- ○ Unemployed
- ○ Retired

4. What is your level of vision impairment

- Sighted
- Vision Impaired
- Totally Blind

5. How many hours per week do you watch television?

- More than 15 hours of television per week
- 11-15 hours of television per week
- 6-10 hours of television per week
- 1-5 hours of television per week
- I do not watch any television

6. How often do you go to the cinema?

- More than once a week
- Once a week
- Once a month
- Once a year
- Never

7. When watching television or watching a movie at home what is your current access strategy?

- Always with video description
- Usually with video description
- Sometimes with video description
- Usually without video description
- Always without video description
- I do not watch television

8. When going to the cinema what is your current access strategy?

- ○ Always with video description
- ○ Usually with video description
- ○ Sometimes with video description
- ○ Usually without video description
- ○ Always without video description
- ○ I do not go to the cinema

9. Please tell us your preferences for video description style.

- ☐ I like it when the video description is distinct from the dialog
- ☐ I like it when the video description is blended with the rest of the sounds and dialog
- ☐ I do not like it when video description is over other sounds
- ☐ I do not like it when video description is over dialog
- ☐ I do not like to hear sounds in the background
- ☐ I like video descriptions at the beginning of the content that gives me extra information about the context
- ☐ I like the credits being described.
- ☐ I prefer a female voice
- ☐ I prefer a male voice

10. What is your level of familiarity with the first show, *The Daily Show* you are about to describe?

- ○ very familiar
- ○ familiar
- ○ somewhat familiar
- ○ not very familiar
- ○ I have never seen this show

11. What is your is your opinion of The Daily Show?

- I really like this show
- I like this show
- I neither like nor dislike this show
- I dislike this show
- I really dislike this show

12. What are the positive aspects of video description in television or in movies?

13. What are the improvements that you would like to have with video description in television or in movies?

Appendix E: Phase 2 Reviewer Review Screen

The survey for this clip is complete.

You are about to listen to some description.

Please make sure your system audio is configured to play audio.

When ready please click this link to begin reviewing the next clip.

You are now listening to clip 1

This is clip 1 of 5 and is approximately **3:00**. You are 20% complete. minutes in length, when it is complete, please click the link **Begin Survey** below to enter your comments.

Once you have reviewed this audio completely please click "Begin Survey" below to enter your remarks.

Begin Survey

Appendix F: Phase 2 Reviewer Clip Review Questionnaire

The purpose of this survey is to capture your opinion on the description you just heard.

1. Rate the overall quality of the video description you just heard?

- ○ Very good
- ○ Good
- ○ Ok
- ○ Poor
- ○ Very Poor

2. Did you feel the description added more humour to the clip?

- ○ Yes a lot
- ○ Yes some
- ○ Did not add or take away
- ○ Took some away
- ○ No, a lot was taken away

3. Did you feel the description added more information to the clip?

- ○ Yes a lot
- ○ Yes some
- ○ Did not add or take away
- ○ Took some away
- ○ No, a lot was taken away

4. Did you feel the description added to the overall entertainment value of the clip?

- Yes a lot
- Yes some
- Did not add or take away
- Took some away
- No, a lot was taken away

5. Please rate the describer's vocabulary level.

- Very good
- Good
- Ok
- Poor
- Very Poor

6. Please rate the describer's style of delivery.

- Very good
- Good
- Ok
- Poor
- Very Poor

7. Please rate the quality of the audio.

- Very good
- Good
- Ok
- Poor
- Very Poor

8. Would you listen to more description from this describer?

- Yes I would definitely
- I probably would
- Neutral
- No I would probably not
- No I would definitely not

9. How would you rate this description overall as compared to professional description you hear on television or at the cinema?

- A lot better
- Better
- As good
- Worse
- A lot worse

10. What are the positive aspects of video description you just heard?

11. What are the negative aspects of video description you just heard?

Appendix G: Ethics Approval

To: Carmen Branje
Re: REB 2007-164: LiveDescribe – Can Amateurs Create Video Description? Phase II
Date: September 7, 2007

Dear Carmen Branje,

The review of your protocol REB File REB 2007-164 is now complete. The project has been approved for a one year period. Please note that before proceeding with your project, compliance with other required University approvals/certifications, institutional requirements, or governmental authorizations may be required.

This approval may be extended after one year upon request. Please be advised that if the project is not renewed, approval will expire and no more research involving humans may take place. If this is a funded project, access to research funds may also be affected.

Please note that REB approval policies require that you adhere strictly to the protocol as last reviewed by the REB and that any modifications must be approved by the Board before they can be implemented. Adverse or unexpected events must be reported to the REB as soon as possible with an indication from the Principal Investigator as to how, in the view of the Principal Investigator, these events affect the continuation of the protocol.

Finally, if research subjects are in the care of a health facility, at a school, or other institution or community organization, it is the responsibility of the Principal Investigator to ensure that the ethical guidelines and approvals of those facilities or institutions are obtained and filed with the REB prior to the initiation of any research.

Please quote your REB file number (REB 2007-164) on future correspondence. Congratulations and best of luck in conducting your research.

Nancy Walton, Ph.D.
Chair, Research Ethics Board

Ec-11-24